On the Use of Phosphorus for Nucleic Acid Structure Determination

Dissertation
zur
Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)
vorgelegt der
Mathematisch-naturwissenschaftlichen Fakultät
der
Universität Zürich
von

Guanya Peng
aus der Volksrepublik China

Promotionskomitee

Prof. Dr. Roland Sigel
Prof. Dr. Anthony Linden
Dr. Vincent Olieric
Dr. Meitian Wang
Dr. Philippe Dumas

Zürich, 2014

Contents

Abstract

With increasing number of nucleic acids sequences, there is a need for a quick and general strategy for nucleic acid structure determination. Anomalous signals from phosphorus intrinsically present in nucleic acid can be used to determine nucleic acid structures via the method of phosphorus single wavelength anomalous dispersion (P-SAD).[1]

As the anomalous signal of phosphorus is very weak, high multiplicity data are required for the P-SAD method. Highly redundant data were collected on RNA and protein-DNA complex crystals. An optimal data collection strategy was found based on systematic analysis of experimental parameters such as wavelength and multiplicity. Data resolution was confirmed to be crucial for the success of phosphorus sub-structure determination. Three targets including two novel structures were successfully determined by P-SAD showing the applicability of the method for small to medium sized RNA. Analysis of the ratio between the number of reflexions and the number of phosphorus demonstrated the potential of applying P-SAD in structure determination of protein-nucleic acid complex.

To extend the method for solving larger RNA structures (>30 nucleotides), Molecular Replacement (MR) with small model fragments[2] can be used to assist the structure determination. Two different MR approaches were studied. The first approach used model fragments in MR search. The second one used MR in combination with anomalous signals. Preliminary studies demonstrated the possibility of using the two approaches to determine a GIR1 branching ribozyme (GIR1) structure containing 188 nucleotides.

---

[1] Z. Dauter and D. A. Adamiak. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **57**: 990–995, 2001.

[2] M. P. Robertson and W. G. Scott. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **D64**: 738–744, 2008.

Zusammenfassung

Mit der wachsenden Anzahl Nukleinsäuresequenzen wächst der Bedarf für eine schnelle und allgemeine Strategie für die Nukleinsäurestrukturaufklärung. Anomale Signale der in Nukleinsäuren enthaltenen Phosphoratome können verwendet werden, um Nukleinsäurestrukturen mit mittels phosphorus single wavelength anomalous dispersion (P-SAD) zu bestimmen.[1]

Da das anomale Signal von Phosphor sehr schwach ist, ist eine hohe Multiplizität der Daten für die P-SAD Methode erforderlich. Dazu wurden hoch redundante Daten an Kristallen von RNA- und Protein-DNA-Komplexen gesammelt. Eine optimale Strategie zur Datensammlung wurde durch systematische Analyse experimenteller Parameter wie Wellenlänge und Multiplizität ermittelt. Es wurde bestätigt, dass die Auflösung der Beugungsdaten entscheidend für die Bestimmung der Phosphor-Substruktur ist. Drei Strukturen, darunter zwei bisher unbekannte, wurden erfolgreich mit P-SAD bestimmt, und somit die Anwendbarkeit der Methode für kleine bis mittelgroße RNA bestätigt. Eine Analyse des Verhältnisses zwischen der Anzahl an Reflexen und der Anzahl der Phosphoratome zeigt das Potential der P-SAD-Methode zur Strukturbestimmung von Protein-Nukleinsäure-Komplexen.

Um das Verfahren auf die Bestimmung von größeren RNA-Strukturen (> 30 Nukleotide) zu erweitern, kann Molekularer Ersatz (MR) zur Unterstützung der Strukturaufklärung verwendet werden.[2] Zwei verschiedene MR-Ansätze wurden untersucht. Zunächst wurden lediglich Modell-Fragmente in der MR-Suche verwendet. Im zweiten Ansatz wurde MR in Kombination mit dem anomalen Signal verwendet. Vorläufige Ergebnisse zeigen, dass beide Ansätze verwendet werden können, um die Struktur eines GIR1 branching ribozyme (GIR1) mit 188 Nukleotiden zu bestimmen.

Part I

Introduction

Chapter I.1

Nucleic Acid Structural Biology

I.1.1 Biological background

Nucleic acids are large biological molecules presented in all living cells and viruses. They play a key role in the storage and replication of hereditary information as well as in the expression of this information through protein synthesis. The two chief types of nucleic acid, DNA (deoxyribonucleic acid) and RNA (ribonucleic acid), are linear polymers with their monomeric units encoding the sequence of hereditary information. The flow of biological information involving DNA and RNA in living organisms is described in the central dogma (Figure I.1.1).[3,4] Gene expression relies on the correct replication of the nucleic acid sequence and transcription of genetic information through interactions of nucleic acid and protein molecules and is fundamental for all cellular functions and heredity.

Nucleic acid is composed of phosphate, sugar and nitrogenous bases. Nitrogenous bases constructing the nucleotides are adenine (A), guanine (G), thymine (T), cytosine (C), and uracil (U). Thymine is found in DNA while in RNA it is substituted by uracil (Figure I.1.2). In DNA structure, the unique set of hydrogen bound on each base allows them to form base pairs and a double-stranded structure. RNA is more often in single-strand form and thus exhibits more possibilities in folding and abundant structural varieties.

---

[3] F. H. C. Crick. *Symp. Soc. Exp. Biol.*, **XII**: 139–163, 1956.
[4] F. H. C. Crick. *Nature*, **227**: 561–563, 1970.

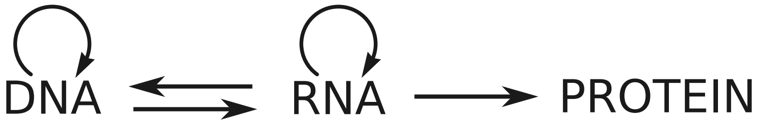


Figure I.1.1: Central dogma of molecular biology. The diagram was used to illustrate transfers of genetic information.

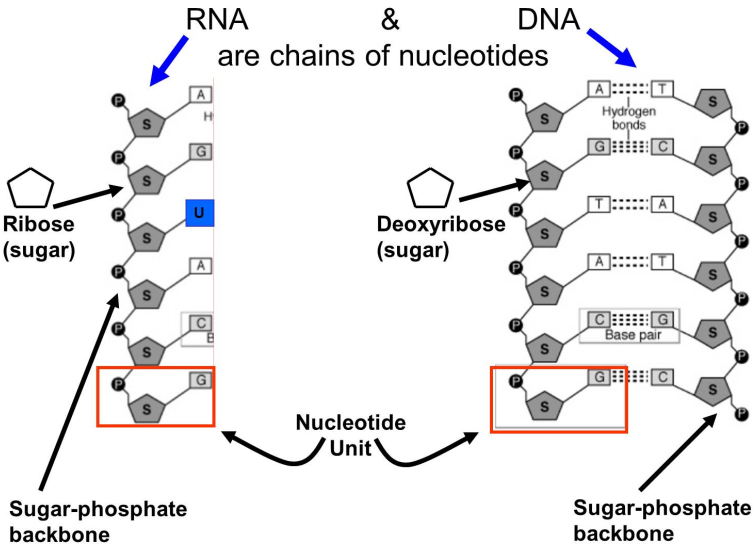
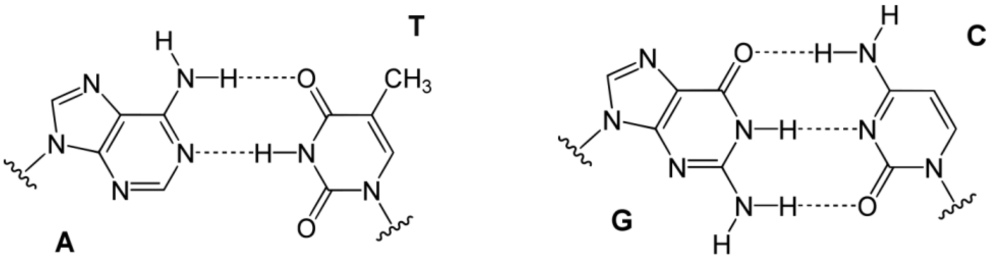


Figure I.1.2: Structural components and monomeric units in RNA and DNA

I.1.2 Nucleic acid structures

I.1.2.1 DNA and RNA structures

DNA frequently exists as a double-stranded molecule.[5] The two DNA strands formed into a double helix holding together by the interaction of the nucleobases in between of the two strands. Each of the four bases in DNA has a unique set of hydrogen bond donors and acceptors that enables it to form base pairs with the other bases. The A-T base pair is one type of the base pairs which is constituted by an adenine and a thymine interacted through two hydrogen bonds. The other type is the G-C base pair which is formed with guanine-cytidine interaction with three hydrogen bonds. These interactions are called Watson-Crick base pairs which are typical for DNA in the double helix (Figure I.1.3). The structure of DNA is stabilized not only by the interaction of the interaction of base pairs between the complementary strands, but also the stacking between the adjacent base pairs. The DNA thus generally adopts a limited number of structural conformations.



(a) **(b)**

Figure I.1.3: The Watson-Crick base pairs. (a) A-T (adenine-thymine) base (b) G-C (guanine-cytidine) base

RNA distinguishes itself from DNA by its sensitivity towards alkaline caused by an additional OH-group on the ribose. RNA shares three bases with DNA: adenine, cytosine and guanine. Uracil is unique to RNA whereas thymine is generally present in DNA. In addition to normal Watson-Crick base pairs, RNA involves non-canonical base pairs defined by the interaction along Hoogsteen edge or Sugar

---

[5] J. D. Watson and F. H. C. Crick. *Nature*, **171**: 737–738, 1953.

edge of a nucleobase (Figure I.1.4).[6] This leads to a rich selection of multiple base interactions in RNA and allows RNA to form more complicated tertiary structures than DNA. The redundant hydrogen donors or acceptors in a RNA base allow interaction with additional nucleotides forming a triples or quadruplex. These formations are important in maintaining the tertiary structure of RNA molecules and are commonly found in RNA structures.

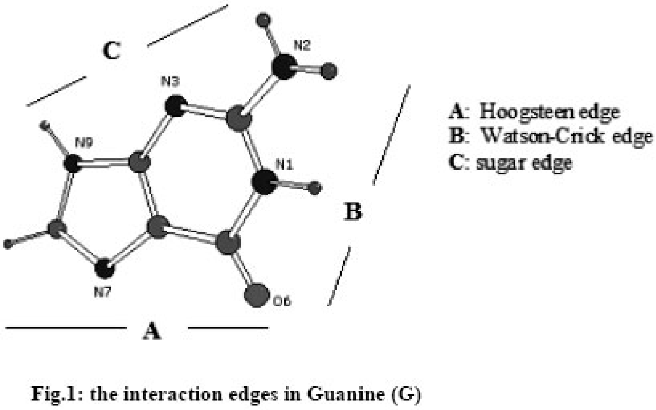


Figure I.1.4: The three interaction edge of a guanine base (A: Hoogsteen, B: Watson-Crick and C: Sugar). Not only the Watson-Crick edge, but also the Hoogsteen edge and sugar edge are involved in forming the base pairs in RNA structures.

I.1.2.2 RNA structural motifs

Structural motifs are locally stable formations sharing similar structural features. They can be considered as the building blocks which are combinatorially linked and constrained by tertiary interactions to create a 3D structure.[7] Compared to DNA, RNA has more abundant base interactions leading to a rich selection of structural motifs. RNA structural motif is defined based on the RNA secondary constrains and its 3D descriptions, backbone conformation, all hydrogen-bonding and base-stacking interactions, and sequence preferences.[7] In 3-dimensional space, a RNA structural motif is a combination of base juxtapositions naturally occurring in the RNA structures or an array of non-Watson-Crick base pairs forming distinc-

---

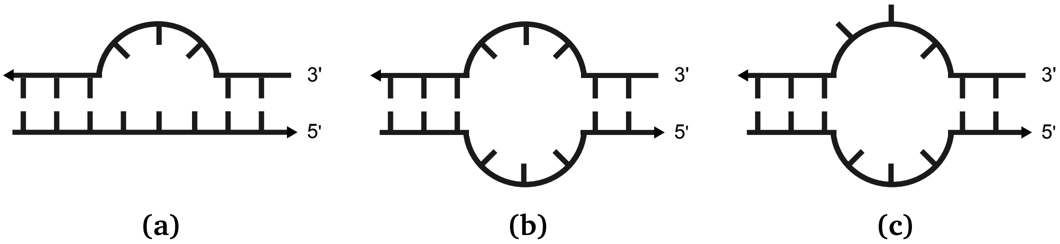
[6] K. Hoogsteen. *Acta Crystallogr.*, **16**: 907–916, 1963.

[7] D. K. Hendrix et al. *Q. Rev. Biophys.*, **38**: 221–243, 2005.

tive folding.[8] Various RNA motifs have been studied. The following discussion will cover several main classes of them.

Hairpin loops Hairpin loop is the structural motif connecting the 3' and 5'ends of a double helix. Hairpin loops are typically closed by a Watson-Crick pairing. Among various kinds of hairpin loops, a four-base hairpin loop motif commonly found in RNA secondary structure is called tetraloop.[9] GNRA loop motif is one example of tetraloop which is especially common in ribosomal RNA. The GNRA tetraloop forms a folding of one guanine in the 5' stack and one adenine in the 3' stack. The fold is stabilized by a G-A base pair and closed by a C-G Waltson-Crick base pair. Other hairpin loop motifs include the T-loop and D-loop motifs of tRNA, the lonepair triloop, and the sarcin-ricin loop.[7]

Internal loops An internal loop adjoins two segments of double helical RNA by inclusion of non-canonical base pairs in at least one strand of the duplex (Figure I.1.5). Internal loops are often binding sites of proteins and are stabilized with tightly bonded water molecules.[7] Various types of internal loops are classified based on base-pairs formation of the adjoining and the stacking of the bases in the loops.[10] When the adjoining non-canonical base pairs inset a bending in a helical structure, the motif can be also called a Bulge.



(a) (b) (c)

Figure I.1.5: Examples of RNA internal loop motifs. (a) A three-base bulge. (b) A metric internal loop with looped-in bases. (c) An internal asymmetric loop with looped-out bases

Junctions Junction loops are formed by the intersection of three or more double helices. These double helices are separated by single-strand sequences of zero or

---

[8] P. B. Moore. *Annu. Rev. Biochem.*, **68**: 287–300, 1999.

[9] C. Woese et al. *Proc. Natl. Acad. Sci. U. S. A.*, **87**: 8467–8471, 1990.

[10] A. Liljas. *Textbook Of Structural Biology*. World Scientific Publishing Company, Incorporated, 2009. 78–88

more residues. Among all junction motifs, the three-way and four-way junctions have been particularly studied and described (Figure I.1.6). Common examples of junction loops are those in tRNA and the hammerhead ribozyme. Many common conformations of junctions can be found in tRNA and hammerhead ribozymes. Most of the junction motifs contain coaxial stacking of helices. The tendency for pairwise coaxial stacking of helical arms, the importance of metal ion interactions in the induction of tertiary folding, and the importance of hairpin or internal loop-loop interactions in stabilization of the tertiary structure are prominent features of the junction loop architectures.[7]

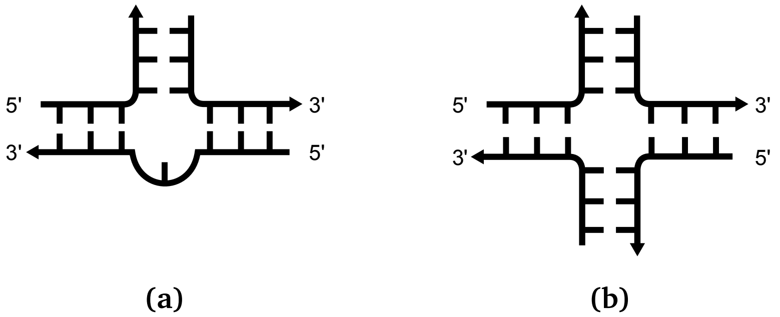


Figure I.1.6: The schematic plot of (a) 3 and (b) 4-way RNA junctions

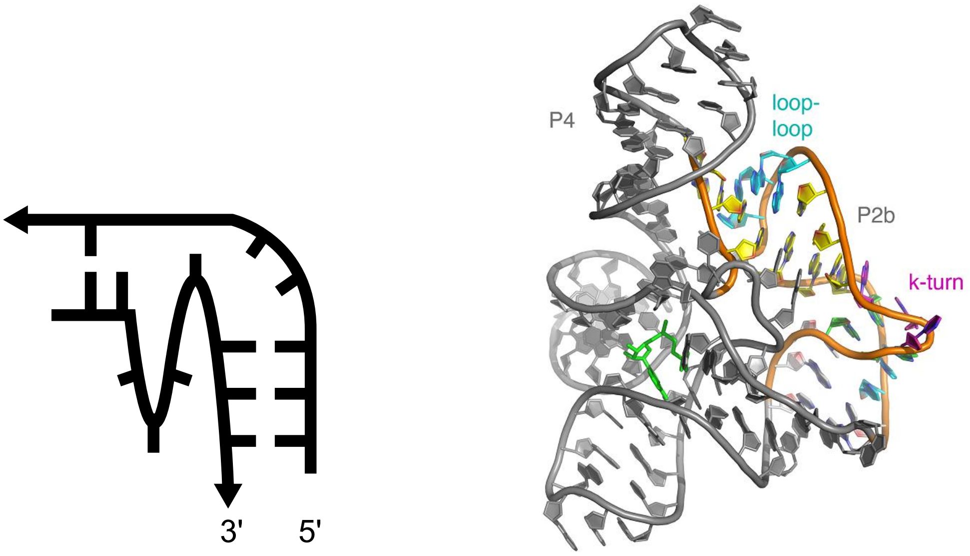
The kink-turn (k-turn) The kink-turn (k-turn) is a two-stranded, loop-helix motif containing 15 residues.[11] It is an asymmetric internal loop characterized by a kink in the sugar-phosphate backbone that causes a sharp turn in the RNA helix.[10] The two separate helices linked by the bending part are a C-stem composed of only Watson-Crick base pairs and a NC-stem of non-canonical helix (Figure I.1.7a). In some large RNA, the kink-turn takes an important role to introduce an abrupt transition in helical trajectory stabilizing the global fold and the whole architecture of the RNA (Figure I.1.7b).[12]

Given the sequence and the secondary structure of the nucleic acid, the misfits can be recognized and stored as library. Considering the high structural similarity of each type of motif and the large amount of RNA motifs in determined structures, the RNA structural motifs can be used as building blocks for solving nucleic acid structures.

---

[11] D. J. Klein et al. *EMBO J.*, **20**: 4214–4221, 2001.

[12] K. T. Schroeder et al. *Structure*, **19**: 1233–1240, 2011.



(a) **(b)**

Figure I.1.7: (a) Schematic of the secondary structure of a k-turn motif (b) 3D structure of a k-turn motif contained in the SAM-I riboswitch[12]

I.1.2.3 Historical review of nucleic acid structures

RNA adopts a wide range of structural diversity to implement its distinct biological functions. The number of deposited entries of nucleic acid structures in Protein Data Bank[13] kept growing since the last 20 years showing the increasing interests of the structural studies of nucleic acid (Figure I.1.8).

Abundant researches have been performed on structural studies of nucleic acid since the very early age of macromolecular crystallography. The three-dimensional double helix structure of deoxyribonucleic acid (DNA) was first elucidated in 1953 by Francis Crick and James D. Watson.[5] The realization that the structure of DNA is in a double-helix formation revealed the relation between the structural properties of the DNA molecule and its mechanism in genetic information storage and transforming and is fundamental to understand the hereditary characters of all living systems.

tRNA$^{\mathbf{phe}}$ Transfer RNA plays an essential role in every step of protein synthesis.

---

[13] H. M. Berman et al. *Nucleic Acids Res.*, **28**: 235–242, 2000.

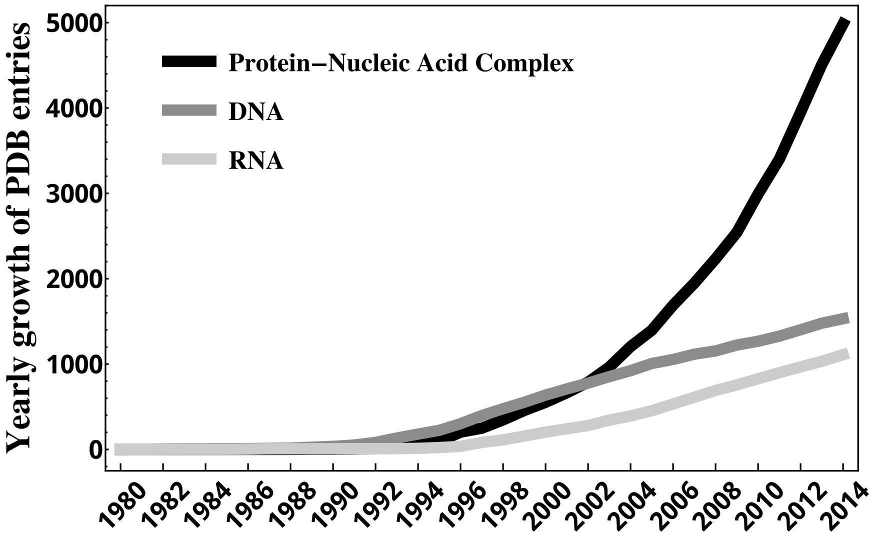


Figure I.1.8: Yearly growth entries of nucleic acid structures in Protein Data Bank since 1980

One of the most remarkable achievements in structural biology in the 1960s and 1970s was the determination of the crystal structure of yeast phenylalanine transfer RNA (tRNA^{phe}). The tRNA^{phe} structure depicted the first solution of a long-chain nucleic acid structure. tRNA^{phe} exhibited various of the tertiary interactions widely appeared in RNA structures establishing a basis for further researches in RNA structures.

tRNA molecule was first purified in 1965. Its sequence demonstrated a cloverleaf structure based mainly on the characteristics of the molecule at specific domains to form stem loop.[14] Based on the work of the purified sequence, further studies were performed on purification and isolation tRNA for crystallographic study. In 1971, crystals of yeast tRNA^{phe} diffracted to 2-3 Å were obtained with a naturally occurring polyamine, which bounded to and stabilized the tRNA.[15] High quality density map was not obtained until the use of heavy metal derivatives in numerous pioneering work. In 1973, a 4 Å map of the tRNA molecule was finally generated to trace the complete backbone unambiguously.[16] Following this solution, numerous investigations on the structure refinement and validation were carried out to give more details structural information demonstrating the pairing

---

[14] R. W. Holley et al. *Science*, **147**: 1462–1465, 1965.

[15] S. H. Kim et al. *Proc. Natl. Acad. Sci. U. S. A.*, **68**: 841–845, 1971.

[16] S. H. Kim et al. *Science*, **179**: 285–288, 1973.

and stacking interactions of the tRNA^{phe}.

Nucleosome structure A nucleosome is the basic unit of DNA packaging in eukaryotes, consisting of a segment of DNA wound in sequence around eight histone protein cores.[17] This structure is analogous to a thread encased around a cylinder.[18] Nucleosomes constitute the fundamental repeating units of eukaryotic chromatin,[19] which is used to pack the large eukaryotic genomes into the nucleus while still ensuring appropriate access to it. Structural studies on nucleosomes were initiated since 1980s. The first evidence of an octamer of histone proteins wraps DNA around itself in about two turns of a left-handed superhelix was proposed in 1984.[20] In 1997, the first near atomic resolution crystal structure of the nucleosome was determined, demonstrating the detailed structural information of the molecule.[21] At the same time, the human alpha-satellite palindromic DNA used to achieve the nucleosome crystal structure was developed.[22,23] Since then, various types of nucleosome core particles have been determined including variants containing histones for different kinds of species. The core structure of nucleosome is highly conserved. Changes of residues between frog and yeast histones can hardly affect the overall structure .[24]

Ribozyme The discovery of ribozyme[25] demonstrated the importance of nucleic acid in the catalytic process in a biological system. Because of the complex enzymatic functions performed by those nucleic acid, their structure can be complicated and diverse in order to achieve their specific functions. Recently, various new catalytic RNA structures were determined indicating its functional characteristics, *e.g.*, structural studies of a hammerhead ribozyme, which illustrated the mechanism of its self-cleaving activity and its influences on gene expression.[26]

---

[17] N. A. Campbell and J. B. Reece. *Biology*. Pearson, Benjamin Cummings, 2005. 360–365

[18] S. Tarakhovsky *Backstage with a command performer* 2003

[19] B. Alberts et al. *Molecular Biology of the Cell 4th Edition: International Student Edition*. Routledge, 2002. 207

[20] T. J. Richmond et al. *Nature*, **311**: 532–537, 1984.

[21] K. Luger et al. *Nature*, **389**: 251–260, 1997.

[22] E. L. Palmer et al. *Anal. Biochem.*, **231**: 109–114, 1995.

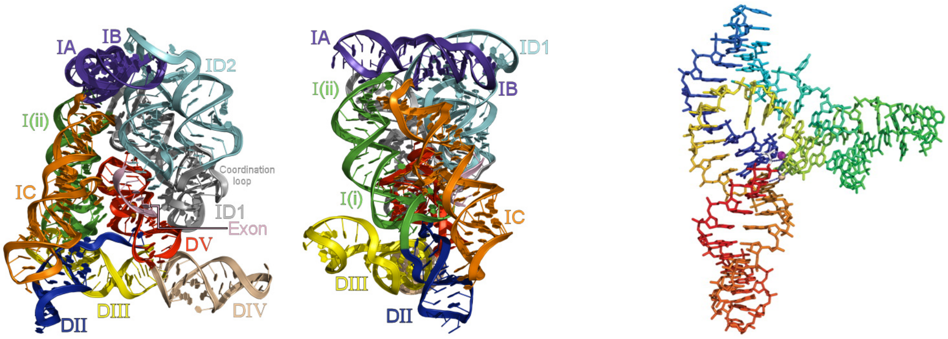
[23] J. M. Harp et al. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **52**: 283–288, 1996.

[24] C. L. White et al. *EMBO J.*, **20**: 5207–5218, 2001.

[25] K. Kruger et al. *Cell*, **31**: 147–157, 1982.

[26] M. Martick et al. *Nature*, **454**: 899–902, 2008.

The three-dimensional structure of a group II intron was recently determined (Figure I.1.9a)[27] allowing visualization of the active site and the intricate network of tertiary interactions that stabilize the molecule. A network of specific structural interactions that promote catalysis of phosphodiester bond formation was found with the investigation of a L1 ligase ribozyme structure (Figure I.1.9b).[28] The structure showed the capability of L1 ligase ribozyme to evolve into a structure capable of catalyzing regiospecific phosphodiester bond ligation.



(a) **(b)**

Figure I.1.9: Schematic for structures of (a) Group II intron active sites[27] and (b) L1 ligase ribozyme[28]

Riboswitch A riboswitch is a regulatory segment of a messenger RNA molecule that binds a small molecule, resulting in a change in production of the proteins encoded by the mRNA.[29–32] The first comprehensive proofs of multiple classes of riboswitches including protein-free binding assays and metabolite-binding riboswitches were discovered in 2002.[33–36].

Many of the earliest riboswitches to be discovered corresponded to conserved sequence motifs in 5' UTRs that appeared to correspond to a structured RNA. Several categories of riboswitches including S-Box or SAM-I riboswitch,[37] the

---

[27] K. S. Keating et al. *RNA*, **16**: 1–9, 2010.
[28] M. P. Robertson and W. G. Scott. *Science*, **315**: 1549–1553, 2007.
[29] E. Nudler and A. S. Mironov. *Trends Biochem. Sci.*, **29**: 11–17, 2004.
[30] A. G. Vitreschak et al. *Trends Genet.*, **20**: 44–50, 2004.
[31] R. T. Batey. *Curr. Opin. Struct. Biol.*, **16**: 299–306, 2006.
[32] B. J. Tucker and R. R. Breaker. *Curr. Opin. Struct. Biol.*, **15**: 342–348, 2005.
[33] A. S. Mironov et al. *Cell*, **111**: 747–756, 2002.
[34] A. Nahvi et al. *Chem. Biol.*, **9**: 1043–1049, 2002.
[35] W. C. Winkler et al. *Proc. Natl. Acad. Sci. U. S. A.*, **99**: 15908–15913, 2002.
[36] W. Winkler et al. *Nature*, **419**: 952–956, 2002.
[37] F. J. Grundy and T. M. Henkin. *Mol. Microbiol.*, **30**: 737–749, 1998.

THI-box in TPP riboswitch,[38] the RFN element or FMN riboswitch[39] and the B12-box in cobalamin riboswitch[40] were observed in gene regulation. Structures of glmS, glycine and PreQ1-I riboswitches were determined recently based on the prediction of homologous UTRs in *Bacillus subtilis* with Bioinformatics. Further development in comparative genomics using additional taxa of bacteria allowed improved algorithms to identify more riboswitch motifs and structures of S-adenosylmethionine riboswitches were determined experimentally.[41–43]

Ribosome The ribosome (from ribonucleic acid and the Greek soma, meaning "body") is a large and complex molecular machine, found within all living cells, that serves as the primary site of biological protein synthesis (translation). The overall molecular structure of the ribosome has been studied since the early 1970s. High resolutions structures were first acquried in 2000s, at resolution of a 5 - 6 Å. The earliest results presenting the structure of the ribosome at atomic resolution were announced in late 2000. The structure of 50S (large prokaryotic) sub-domain was determined from the bacterium *Deinococcus radiodurans*[44] and the archaeon *Haloarcula marismortui*.[45] The 30S sub-domain was determined from *Thermus thermophilus*.[46] In 2001, these structures were supported to rebuild the entire *T. thermophilus* 70S fragment at a resolution of 5.5 Å.

More and more structural studies were performed with larger complex nucleic acid revealing their distinct functions in biological systems. The increasing number of nucleic acid structures indicated that there is a need for a fast and generalized method for structure determination.

---

[38] J. Miranda-Ríos et al. *Proc. Natl. Acad. Sci. U. S. A.*, **98**: 9736–9741, 2001.
[39] M. S. Gelfand et al. *Trends Genet.*, **15**: 439–442, 1999.
[40] C. V. Franklund and R. J. Kadner. *J. Bacteriol.*, **179**: 4039–4042, 1997.
[41] J. E. Barrick et al. *Proc. Natl. Acad. Sci. U. S. A.*, **101**: 6421–6426, 2004.
[42] Z. Weinberg et al. *Genome Biol.*, **11**: R31, 2010.
[43] Z. Weinberg et al. *Nucleic Acids Res.*, **35**: 4809–4819, 2007.
[44] F. Schluenzen et al. *Cell*, **102**: 615–623, 2000.
[45] N. Ban et al. *Science*, **289**: 905–920, 2000.
[46] B. T. Wimberly et al. *Nature*, **407**: 327–339, 2000.

I.1.3 Nucleic acid crystallography

The most common method to solve nucleic acid structures is X-ray crystallography. It represents more than 70% of the depositions in the Protein Data Bank[13] (Figure I.1.10).

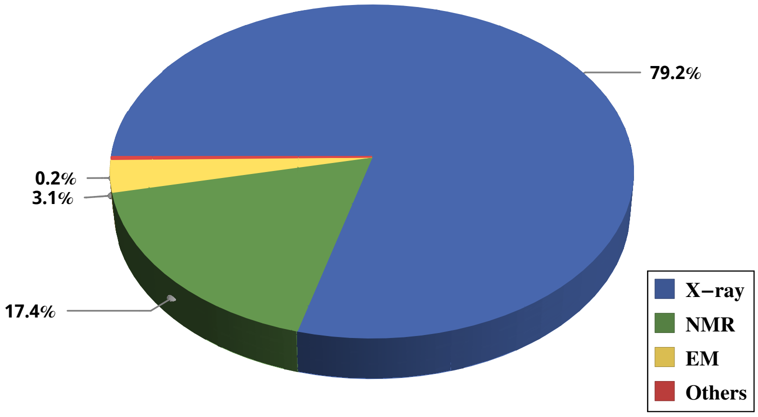


Figure I.1.10: Methods for solving the macromolecules containing nucleic acid in Protein Data Bank

The crystalline phase appears in nucleic acid molecules solution when a certain concentration is reached. Crystallization can be quite challenging, as one needs to find the nucleation zone where both of the crystals and precipitates grow and the competition between them consumes the crystallization resources (Figure I.1.11).[47] Especially for nucleic acid, the large size of the molecules and the fragility of their single-strand backbones increase the difficulty of crystallization. Vapor diffusion is the most commonly used method in the crystallization of macromolecules such as nucleic acid. It required fine adjustment of crystallization condition.

Synchrotron radiation is the most widely used source for measuring X-ray diffraction data. Its high flux, high brilliance and highly collimated photon beam with a broad spectrum are well suited for macromolecular crystallography. In addition, the high stability of the synchrotron source allows very accurate measurement

---

[47] N. E. Chayen and E. Saridakis. *Nat. Methods*, **5**: 147–153, 2008.

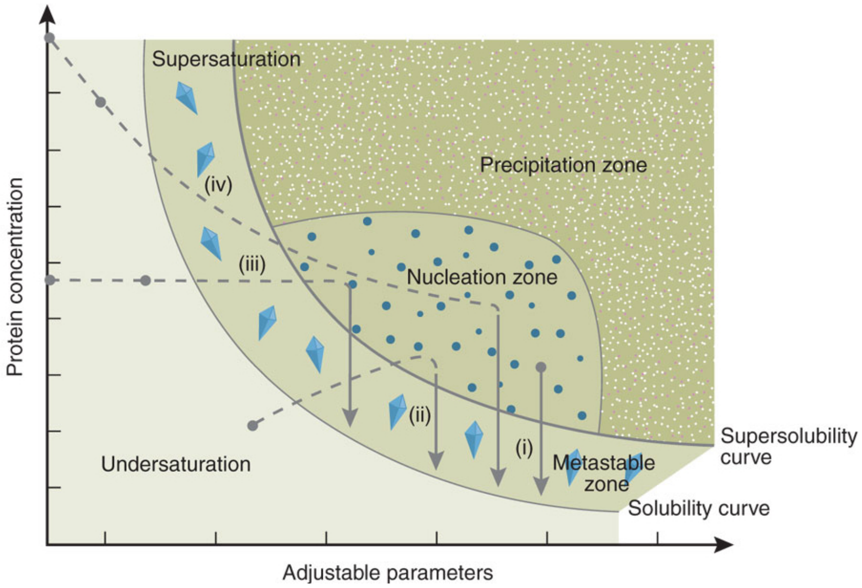


Figure I.1.11: The phase diagram for the macromolecular crocrystallization. In undersaturation and precipitation phases, no crystal growth. Crystal seeds will appear at nuleation zone and get to crystalline state in supersaturation phase where screening should be performed to obtain crystals.[47]

of weak scattering signals. Crystallography data are collected with the rotation method. Due to the compact design of beamline, crystals are commonly rotated along Ω axis. Recent development of multi-axis goniometer allows crystals to be orientated more freely. Advantages of crystal geometry can be taken to obtain optimal data quality for solving nucleic acid structures.

Chapter I.2

Macromolecular Structure Determination

To determine nucleic acid structures with X-ray crystallography, the electron density map needs to be reconstructed from the diffraction experiments. The electron density distribution of a crystalline solid in real space relates to its reciprocal structure by *Fourier transform* (equation D.1.1).

As the phases information (φ) in equation D.1.1 can not be measured directly, structure determination with X-ray diffraction is not a trivial matter and various methods (*e.g.*, direct method, Molecular Replacement, heavy atom isomorphous replacement, isomorphous replacement with anomalous scattering and single/multi-wavelength anomalous dispersion) were developed to retrieve the phase information and determine the macromolecular structures.

This chapter will describe the principle and applicability for each of the phasing methods and introduce the concept of using phosphorus anomalous signals in supplement of the experimental phasing.

I.2.1 Direct methods

Direct methods refer to methods using phases and amplitudes relationships to retrieve phase information from experimental data directly. The basic principle of these methods are based on the restriction that the electron density must be non-negative and atomically distributed. With this assumption, the *Karker-Kasper inequalities* (equation D.2.3) and *Karle-Hauptman tangent formula* (equation D.2.1b) can be derived to estimate phases directly from intensities.[48]

For solving macromolecule such as protein or nucleic acid structures with Direct methods, atomic resolution beyond 1.2 Å is required.[49] This so called *Shleldrick's rule*[50] limits the applicability of the Direct methods for protein and nucleic acid structure. However, the Direct methods are still the powerful tools to solve the heavy atoms sub-structure and are implemented in many structure determination software, *e.g.*, *SnB*,[51] *SHELXD*,[52] *SHARP*,[53] *MOLREP*[54] and *Hyss*.[55] For the sub-structure determinations and analysis in this thesis, *SHELXD* was used as a standard software.

---

[48] J. Karle and H. Hauptman. *Acta Crystallogr.*, **3**: 181–187, 1950.

[49] R. J. Morris and G. Bricogne. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **59**: 615–617, 2003.

[50] G. M. Sheldrick. *Acta Crystallogr. Sect. A Found. Crystallogr.*, **46**: 467–473, 1990.

[51] R. Miller et al. *J. Appl. Crystallogr.*, **27**: 613–621, 1994.

[52] T. R. Schneider and G. M. Sheldrick. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **58**: 1772–1779, 2002.

[53] E. d. L. Fortelle and G. Bricogne. *Methods Enzymol.*, **276**: 472–494, 1997.

[54] A. Vagin and A. Teplyakov. *J. Appl. Crystallogr.*, **30**: 1022–1025, 1997.

[55] R. W. Grosse-Kunstleve and P. D. Adams. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **59**: 1966–1973, 2003.

I.2.2 Molecular Replacement

When a homologous model is available, prior knowledge of the actual geometry as well as atomic connectivity of the structure can be a source of information to facilitate structure determination.[56] Methods involving fitting preconceived model to the observed data to extract useful phase information are named Molecular Replacement (MR).

I.2.2.1 Procedure of Molecular Replacement search

The core of MR is to search the correct orientation and position of the molecule in the unit cell. The searching procedure involves rigid-body placement of a structurally analogous or an identical search model in the asymmetric unit over all conceivable orientations and positions. Each placement is evaluated through the agreement between the calculated and observed structure factors using Patterson method[57] or maximum-likelihood method.[58,59] The best fitting model resembles the target structure.

To find out the correct orientation and position of the search model in the unit-cell, two steps of searching are performed in MR. First the orientation of the search model is determined via rotation search and then translation search is carried out to locate the search model in the unit-cell.[56] While the orientation and location of the search model are determined, the remaining tasks of MR relate to refinement. Maximum-likihood methods are used to generate probability functions to evaluate the fitting between model and experimental data. Parameters such as Log-likelihood gain (LLG) and Z-scores can be used to check the solutions. In some difficult cases when data are poor or search model is incomplete or contains large errors, all solutions may have similar low scores. The best way to judge the

---

[56] M. G. Rossmann and D. M. Blow. *Acta Crystallogr.*, **15**: 24–31, 1962.

[57] A. L. Patterson. *Zeitschrift für Krist. Mater.*, **90**: 517–542, 1935.

[58] V. Y. Lunin and A. G. Urzhumtsev. *Acta Crystallogr. Sect. A Found. Crystallogr.*, **40**: 269–277, 1984.

[59] R. J. Read. *Acta Crystallogr. Sect. A Found. Crystallogr.*, **42**: 140–149, 1986.

quality of MR solution is to check the quality of the electron density map.

I.2.2.2 Molecular Replacement for nucleic acid

MR relies on the knowledge of structural features of the target molecule. Thus search model preparation is crucial. If the homologous model is not available, structural fragment library can be used to facilitate the MR search. The library can be built by fragments extracted from structures in database or predicted from secondary structures. While the library is built, it can be used for *de novo* structure determination with MR. For instance, a 71-nucleotides L1 ligase ribozyme nucleic acids structure was shown to be possible to solve by MR with predicted model fragments based on its secondary structure.[2]

There are also some peculiarities when applying MR for solving nucleic acids structures. High internal symmetry of nucleic acid helices can create additional complications in the MR search and may lead to difficulties in the placement of search model.[60] In addition, the success of MR rotation and translation search requires accurate and complete low resolution data.[61]

I.2.3 Heavy Atom Isomorphous Replacement

Heavy atoms isomorphous replacement phasing refers to methods of solving the phase problem with heavy-atom substitution.[62] Method using a single heavy atom in the derivative structure is called single heavy atom isomorphous replacement phasing (SIR). The soaked heavy-atom in the derivative structure provides slight yet measurable intensity differences compared to the native structure. The intensity differences are then used to determine the heavy-atom sub-

---

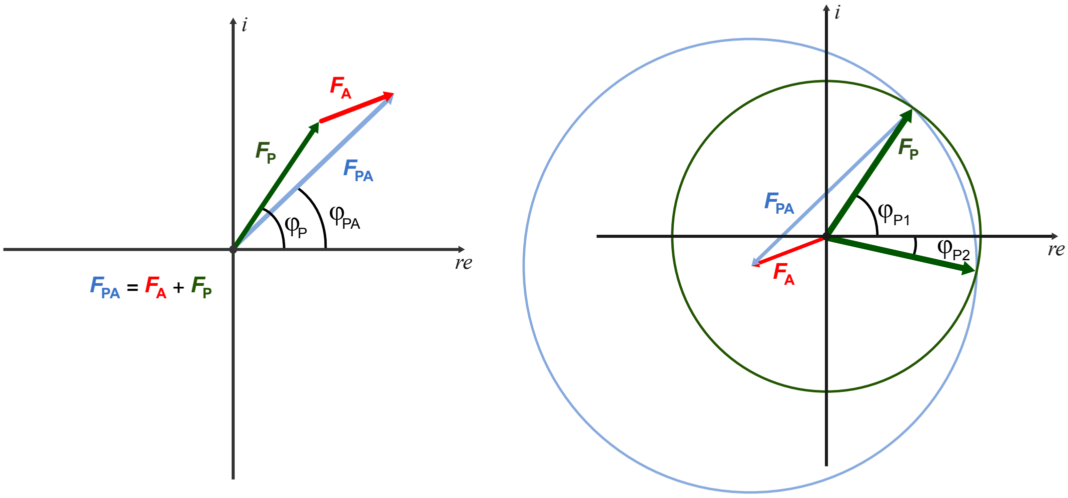
[2] M. P. Robertson and W. G. Scott. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **D64**: 738–744, 2008.

[60] I. Baikalov and R. E. Dickerson. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **54**: 324–333, 1998.

[61] E. Dodson. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **64**: 17–24, 2008.

[62] D. W. Green et al. *Proc. R. Soc. A Math. Phys. Eng. Sci.*, **225**: 287–307, 1954.

structure (Figure I.2.1a).[63] Then the estimated phases for the native structure can be obtained with the known phases of the sub-structure as showed in Figure I.2.1b.



(a) **(b)**

Figure I.2.1: Harker construction[63] for SIR. (a) The angular relation between amplitudes of the native structure (F_P), have-atom derivative (F_{PA}) and the single atom substructure (F_A) which is used in determine the phases of heavy-atoms, (b) Harker construction showing phases of the native structure are determined by the intersection between the circle of the have-atom derivative structure factor and the structure factor of native structure. With SIR only, two possible sets of phases (φ_{P1}, φ_{P2}) are determined

For SIR, two sets of phases are obtained (Figure I.2.1b). Density modification can be utilized to resolve this phase ambiguity. With multi-isomorphous replacement (MIR), the phase ambiguity is automatically solved as, in theory, there is only one intersection in its Harker construction.[63] To apply heavy atom isomorphous replacement in solving nucleic acid structure, heavy atoms must be incorporated into the nucleic acid molecule. The heavy atom binding often causes non-isomorphism between native and derivative crystals. In addition, the heavy atoms may increase X-ray absorption during exposure.

---

[63] D. Harker. *Acta Crystallogr.*, **9**: 1–9, 1956.

I.2.4 Methods using anomalous signals

I.2.4.1 Breaking Friedel's law

The atomic scattering factor $f_{\mathbf{s},\lambda}$ is composed of three parts and can be expressed explicitly as,[64]

$$f_{\mathbf{s},\lambda} = f_{\mathbf{s}}^{0} + f'_{\lambda} + i f''_{\lambda} \tag{I.2.4.1}$$

where $f_{\mathbf{s}}^{0}$ is the wavelength independent but scattering angle dependent term and f'_{λ} and f''_{λ} denotes the wavelength dependent anomalous scattering contributions. f'' has a 90° phase shift added on the dispersive term f'_{λ} and thus break down the equality between the amplitudes of a Bijvoet pair, *i.e.*, *Friedel's law* (Figure I.2.2).[65] This inequality of Bijvoet pairs' amplitudes is a basis for applying experimental phasing using the anomalous scattering.

---

[64] M. Vijayan and S. Ramaseshan. *International Tables for Crystallography*, **B**: 265–269, 2006.

[65] G. Friedel. *Comptes Rendus l'Académie des Sci.*, **157**: 1533–1536, 1913.

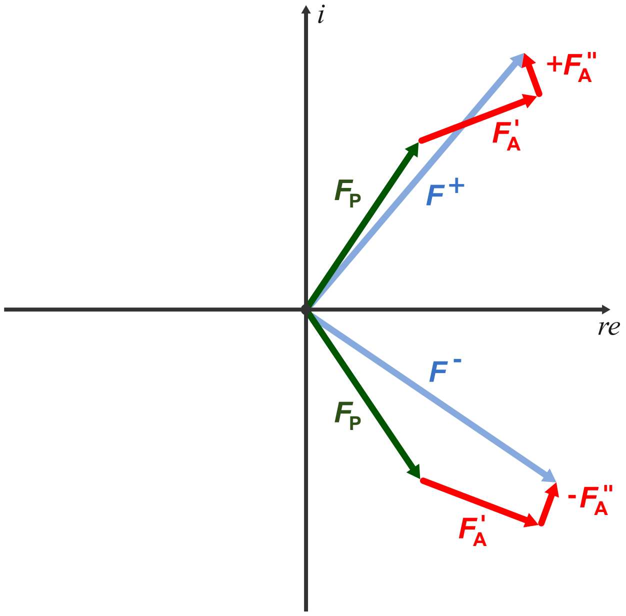


Figure I.2.2: The breakdown of Friedel's law. Friedel's Law states that members of a Friedel pair (centrosymmetric reflections, *hkl* and $\overline{hkl}$) have equal amplitude and opposite phase. The Friedel's law is not valid when an anomalous scatter is present. The imaginary term, f'' in atom scatting factor $f_{\mathbf{s},\lambda} = f_{\mathbf{s}}^0 + f'_\lambda + i f''_\lambda$, represents a 90° phase shift for anomalous contribution and leads to $|F^+| \neq |F^-|$. The difference in the amplitude of a Bijvoet pair, $\Delta F^\pm = |F^+| - |F^-|$, is called the Bijvoet difference or anomalous difference.

I.2.4.2 Isomorphous Replacement with anomalous scattering

The amplitude difference of a Bijvoet pair can be used in Patterson or Direct method to locate the anomalous scattering atoms. The phases of the anomalous scatters and the measured amplitudes of native and anomalous derivative structures can be then used to determine the phases. The breakdown of Friedel's law provides additional information allows selecting only one intersection in the Harker construction to suppress phase ambiguity (Figure I.2.3).

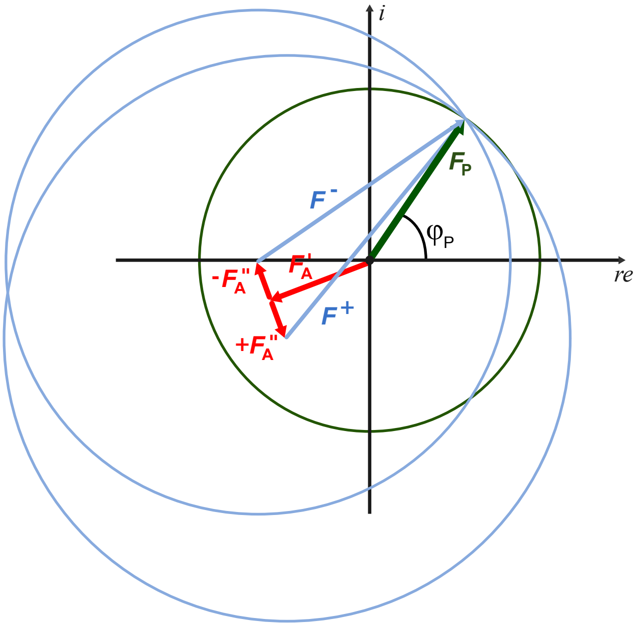


Figure I.2.3: Harker construction for phasing with isomorphous replacement with a single anomalous scatter. The anomalous terms ($\pm F''_{\mathrm{A}}$) of structure factors lead to an anomalous difference and an intersection in the Harker construction showing the determined phases of the native structure.

Commonly used heavy-atom reagents for solving protein structures are shown in Table I.2.1.[66] For nucleic acid, magnesium, zinc, manganese, barium, bromine and iridium are frequently used.[67–69] The preparation of the heavy-atom derivative requires specific experimental efforts with well controlled concentration in the macromolecule solution. Care needs to be taken to protect the fragile nucleic acids crystals, if soaking is performed. This method also suffers from the problem of non-isomorphism that adds additional errors in the determined structures.

---

[66] D. Carvin et al. *International Tables for Crystallography*, **F**: 247–255, 2006.
[67] F. E. Reyes et al. *Methods Enzymol.*, **469**: 119–139, 2009.
[68] J. H. Cate and J. a. Doudna. *Methods Enzymol.*, **317**: 169–80, 2000.
[69] A. Y. Keel et al. *Structure*, **15**: 761–772, 2007.

Table I.2.1: The 23 most commonly used heavy-atom reagents for proteins. The first column shows the number of times the reagents has been used in the analyses included in the heavy-atom data bank.[66]

No.	Compound
287	Potassium tetrachloroplatinum (II)
111	Potassium dicyanoaurate (I)
103	Uranyl acetate
101	Mercury (II) acetate
98	Mecury (II) chloride
85	Ethylmercurythiosalicylate (EMTS)
82	Postassium tetraiodomercurate (II)
81	*para*-Chloromecuriobenzenesulfonate (PCMBS)
75	Trimethyllead (IV) acetate
73	Potassium pentafluorooxyuranate (VI)
73	Phosphatotris (ethylmecury)
61	Potassium tetranitritoplatinum (II)
60	Uranyl nitrate
58	Potassium tetracyanoplatinum (II)
57	Dichlorodiammineplatinum (II)
51	Potassium hexachloroplatinum (IV)
51	Methylmercury chloride
44	Potassium tetrachloroaurate (III)
42	*para*-Chloromercurybenzoate (PCMB)
39	Lead (II) acetate

I.2.4.3 Single/Multi-wavelength anomalous dispersion

Methods of single/multi-wavelength anomalous dispersion (SAD/MAD) can be readily performed to determine the structures of native macromolecules.[70] For SAD, single wavelength is used to deduce the anomalous differences and predict the possible phases. As showed in Figure I.2.4, the two ambiguous phases are symmetrically located about $\varphi_A - 90^\circ$. This phase ambiguity can be resolved by density modification. The multi-wavelength anomalous diffraction (MAD) is performed with data collected at different wavelengths, which allows to solve the phase ambiguity.

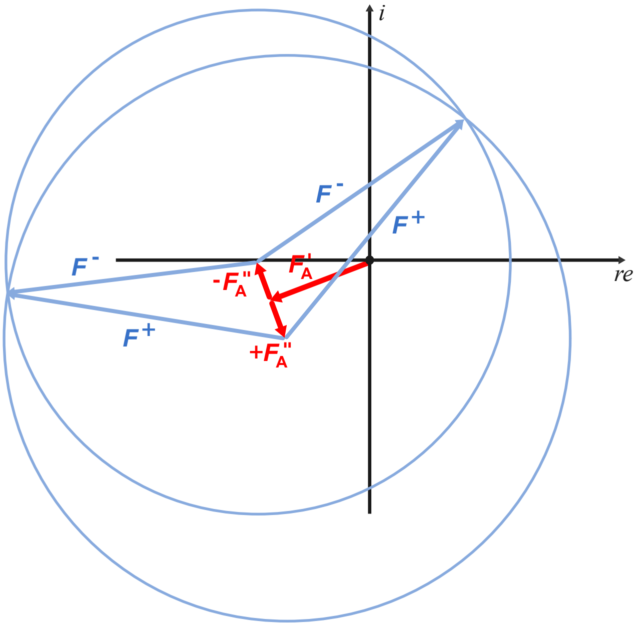


Figure I.2.4: Harker construction for SAD phasing. With a single wavelength, two phases circles are composed by the vector of $+F$ and $-F$ added at the end of vector of $+\mathrm{F}''_{\mathrm{A}}$ and $-\mathrm{F}''_{\mathrm{A}}$ respectively. The two phase circles meet at two points leading to an ambiguous phase determination.

One way to use SAD/MAD is to introduce selenomethionine in place of methionine residues of a protein in order to minimize non-isomorphism.[71,72] Sulfur is naturally presented in sulfur-containing amino acids such as methionine and

---

[70] J. Karle. *Int. J. Quantum Chem.*, **18**: 357–367, 1980.
[71] W. A. Hendrickson et al. *Proc. Natl. Acad. Sci. U. S. A.*, **86**: 2190–2194, 1989.
[72] W. A. Hendrickson et al. *EMBO J.*, **9**: 1665–1672, 1990.

cysteine. Therefore, for SAD/MAD phasing with the anomalous signal of sulfur (S-SAD), no experimental effort is needed to prepare derivative structures and non-isomorphism problem can be avoided. S-SAD was first successfully applied to determine the crambin structure in 1981.[73] The use of the weak anomalous scattering from low Z elements (< 20) has been shown very powerful in the so-called native-SAD method.[74,75] Like sulfur, phosphorus is naturally presented in nucleic acid and therefore has potential to help in nucleic acid structure determination.

As the anomalous contribution is normally a few percent of the scattering contributed by the entire molecule, specific wavelengths at absorption edges have to be chosen to maximize the anomalous differences in SAD or MAD methods. At current synchrotron radiation source, the absorption edges for S/P native-SAD is not attainable. The only way to increase anomalous scattering for S/P native-SAD is to use longer wavelength or decrease the energy. The minimal energy at beamline X06DA/PXIII is 6 keV. However, the absorption dose is higher and the diffraction angles are limited at lower energy. Therefore, compromise between the strength of the anomalous scattering, absorption dose and diffraction angles has to be made to obtain an optimal signal for S/P native-SAD.

---

[73] W. A. Hendrickson and M. M. Teeter. *Nature*, **290**: 107–113, 1981.

[74] Q. Liu et al. *Science*, **336**: 1033–1037, 2012.

[75] Q. Liu et al. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **69**: 1314–1332, 2013.

Chapter I.3

Using Phosphorus Anomalous Signals for Nucleic Acid Structure Determination

As the phosphates are naturally presented in the backbone of nucleic acid structures, there is no non-isomorphism problem for phasing with the phosphorus anomalous signal. With the method of phosphorus single-wavelength anomalous dispersion (P-SAD), the experimental effort of heavy-atoms derivative preparation can be omitted to speed up the nucleic acid structure determination. Thus, automatic pipeline for *de novo* experimental phasing is possible.[76] A structure of a left-handed Z-DNA hexamer duplex d(CGCGCG)$_2$ were successfully determined with the P-SAD method[1] showing its feasibility in determining small sized nucleic acid structures.

I.3.1 Weak anomalous signals

The anomalous signal is evaluated by using the *Bijvoet* ratio which is the ratio between the anomalous difference (ΔF_{anom}) and the measured amplitude (F). Calculation based on simulated data illustrated one disulfide bridge in a protein with 120 amino acids was able to solve the protein structure.[77] This amount of signal is approximately 0.6% of $\langle \Delta F_{\text{anom}}/F \rangle$ and is considered as the minimal signal required for solving protein structures with the SAD method. For nucleic

---

[76] E. Micossi et al. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **58**: 21–28, 2001.

[1] Z. Dauter and D. A. Adamiak. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **57**: 990–995, 2001.

[77] B. C. Wang. *Methods Enzymol.*, **115**: 90–112, 1985.

acid, the estimated anomalous signal $\langle \Delta F_{\mathrm{anom}}/F \rangle$ at wavelength from 1.6 to 2.0 Å is larger than 0.6% (Table I.3.1). It suggests that, in principle, there is a chance to solve the nucleic acid structure with the anomalous signals from the intrinsic phosphorus atoms.

Table I.3.1: The estimated anomalous signal for the nucleic acid at different wavelength calculated with $\langle \Delta F_{\mathrm{ano}}/F \rangle = \sqrt{2}/6.7 f'' \sqrt{N_A/N_P}$, where ΔF_{ano} is the anomalous difference and F is the amplitude of the reflections and f'' is the theoretical value of the imaginary component for the atomic scattering factors of phosphorus and N_A is the number of the anomalous scatters and N_P is the number of the other atoms within the whole molecule. $\langle \Delta F_{\mathrm{ano}}/F \rangle$ is used to estimate the anomalous strength of the overall structure.

λ (Å)	f''	$\langle \Delta F_{\mathrm{ano}}/F \rangle$ (%)
1.6	0.47	2.1
1.8	0.59	2.6
2.0	0.71	3.1

However, similar to sulfur, the anomalous signal of phosphorus is extremely weak in attainable energy range at a synchrotron radiation source (5 ∼ 12.4 keV) (Figure I.3.1).[78–81] The K-edge of phosphorus at 2143 eV is not achievable at classical synchrotron MX beamline. To enhance the anomalous dispersion, long wavelength can be used. Long wavelength synchrotron beamlines (*e.g.*, beamline I23, Diamond[82]) are under constructions towards the goal of native sulfur or phosphorus SAD phasing.

---

[78] D. T. Cromer. *Acta Crystallogr.*, **18**: 17–23, 1965.

[79] D. T. Cromer and J. B. Mann. *Acta Crystallogr. Sect. A Found. Crystallogr.*, **24**: 321–324, 1968.

[80] D. T. Cromer and D. A. Liberman. *J. Chem. Phys.*, **53**: 1891–1898, 1970.

[81] D. T. Cromer and D. A. Liberman. *Acta Crystallogr. Sect. A Found. Crystallogr.*, **37**: 267–268, 1981.

[82] V. Mykhaylyk and A. Wagner. *J. Phys. Conf. Ser.*, **425**: 012010, 2013.

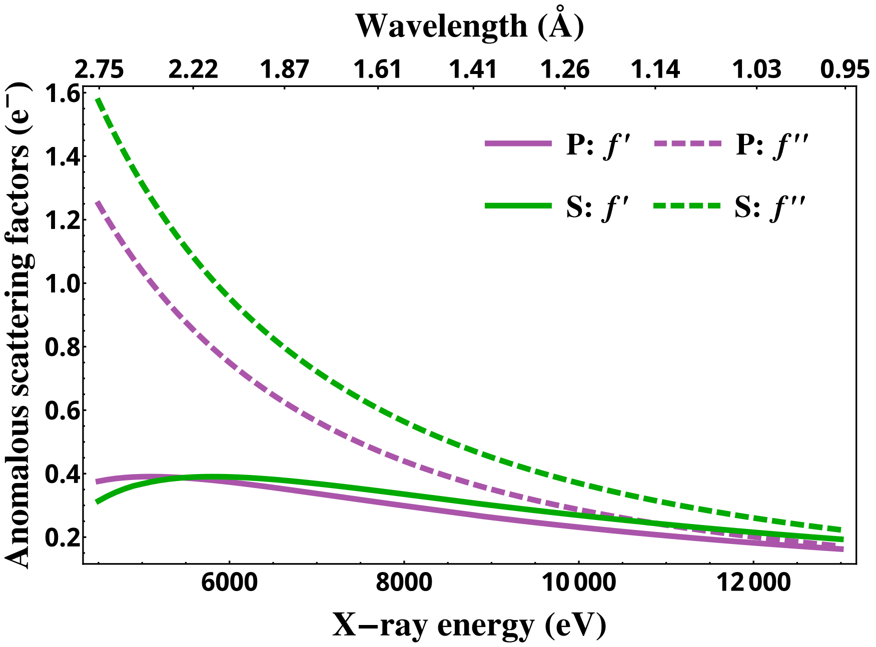


Figure I.3.1: The anomalous scattering coefficients f' and f'' of phosphorus and sulfur atoms

I.3.2 Effect of absorption

As showed in Figure I.3.1, longer wavelength provides more significant anomalous scattering. However, the absorption effect is increased with longer wavelength. The absorption dose introduces radiation damage to molecules that impairs the data quality and restricts the measured resolution. As there is a trade-off between the anomalous strength and absorption, proper wavelength is required to pursue optimal data quality for P-SAD phasing, in which radiation damage is more significant for large crystals (Figure I.3.2).[74] Therefore, small crystals (< 10 μm) matching micron-sized beam are meant to be used for data collected at wavelength > 3.0 Å.[83] Abortion through the beam path is another issue affecting data accuracy. Since the X-ray absorption of helium is by two to three orders lower than air,[84] an helium filled chamber can be used to minimize the adsorption.

---

[74] Q. Liu et al. *Science*, **336**: 1033–1037, 2012.
[83] R. Sanishvili et al. *Proc. Natl. Acad. Sci. U. S. A.*, **108**: 6127–6132, 2011.
[84] K. Djinović Carugo et al. *J. Synchrotron Radiat.*, **12**: 410–419, 2005.

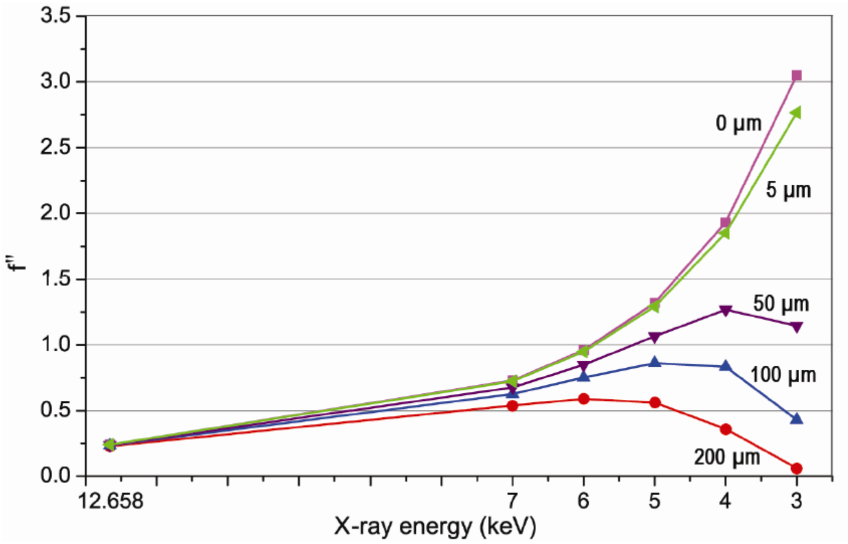


Figure I.3.2: Transmitted anomalous signals as function of X-ray energy through water. X-ray transmission of anomalous signals from sulfur (f'' in electrons) through various thickness of water: red: 200 μ m; blue; 100 μ m; purple: 50 μ m; green: 5 μ m; and magenta: 0 μ m. The transmitted anomalous signals are sample thickness dependent.

I.3.3 MR and anomalous signals

The applicability of P-SAD method is limited by the weak anomalous signals and absorption effect. To improve P-SAD method and extend its potential in solving larger nucleic acid structures, P-SAD will likely benefit from combined approaches. Molecular Replacement (MR), as mentioned in section I.2.2.2, uses the phase information from homologous model and the success of this method does not crucially depend on high resolution data. Thus, MR can be a supplement for SAD or MAD phasing in order to enhance the phasing power. In the other way around, SAD or MAD phasing can provide independent experimental phase information to support the model search or refinement in MR.

So far, some attempts with different approaches have been tried to use anomalous scattering signals in combination with MR in structure determination.

(1) Use MR to determine the anomalous sub-structure for initiation of SAD phasing and then compare the anomalous sub-structures obtained from SAD phasing

and MR in order to check the consistency of phasing results.[85]

(2) Use the anomalous sub-structure obtained from SAD phasing to generate an electron density map which is unbiased by MR search model and use the map to identify the correct solution among a group of weak MR solutions generated from inaccurate search models.[86]

(3) Combine the rotation functions of MR for full structure search in intensities and that of MR for anomalous sub-structure search in anomalous differences. Then generate a new score from this combined rotation functions and use it to evaluate the MR rotation search.[87]

(4) Use anomalous signals to refine the initial MR solution and get rid of model bias with SAD likelihood function.[88,89]

In this thesis, feasibility of incorporating anomalous signals in MR search for solving nucleic acid structures was investigated. Aiming at *de novo* structure determination of nucleic acids structures, the strategy of using nucleic acid structural fragments library to facilitate MR search in either intensities or anomalous differences was studied.

---

[85] J. P. Schuermann and J. J. Tanner. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **59**: 1731–1736, 2003.

[86] E. N. Baker et al. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **51**: 282–289, 1995.

[87] J. Unge et al. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **67**: 729–738, 2011.

[88] A. J. McCoy et al. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **60**: 1220–1228, 2004.

[89] A. J. McCoy et al. *J. Appl. Crystallogr.*, **40**: 658–674, 2007.

Chapter I.4

Goal of Thesis

This thesis intends to examine the feasibility and limitation of using phosphorus anomalous signals in solving nucleic acids structures. To clarify the requirement in each step of structure determination, the thesis will study the effect of various experimental parameters such as wavelength and multiplicity on the performance of the method. Based on these requirements, practical data collection strategy for P-SAD will be investigated. Finally, the goal is to apply the P-SAD method to determine novel structures in order to test its general applicability. To extend the method for solving larger structures, the thesis will also study the possible approach and applicable conditions of using Molecular Replacement in combination with phosphorus anomalous signals.

Part II

Materials and Methods

Chapter II.1

Nucleic Acid Targets

II.1.1 Target molecules

The sarcin ricin loop RNA (SRL) containing 27 nucleotides[90,91] was used as the target to perform systematic studies and test the feasibility of the P-SAD method. A novel structure of an octameric RNA duplex (2 × 8-mer) containing 2 × 8 nucleotides was successfully determined with P/native-SAD method. Two larger structures, *i.e.*, a Dimerization Initiation Site RNA with 2 × 23 nucleotides[92] and a GIR1 branching ribozyme with 188 nucleotides,[93] were also used to test the feasibility of the method in solving larger RNA structures. Novel structure of a σ^{E2}–TGTCAAA protein-DNA complex[94] containing 2 × 7 nucleotides was successfully determined in the course of the thesis illustrating the potential of the method for protein-nucleic acid complex. The schematic plots for all molecules are displayed in Figure II.1.1. Detailed information for each target can be found in Table II.1.1. The subsequent sections will discuss the synthesis, purification and crystallization of the five targets.

---

[90] V. Olieric et al. *RNA*, 707–715, 2009.

[91] C. C. Correll et al. *J. Mol. Biol.*, **292**: 275–287, 1999.

[92] E. Ennifar et al. *Structure*, **7**: 1439–1449, 1999.

[93] M. Meyer et al. *Proc. Natl. Acad. Sci. U. S. A.*, **111**: 7659–7664, 2014.

[94] S. Campagne et al. *Nat. Struct. Mol. Biol.*, **21**: 269–276, 2014.

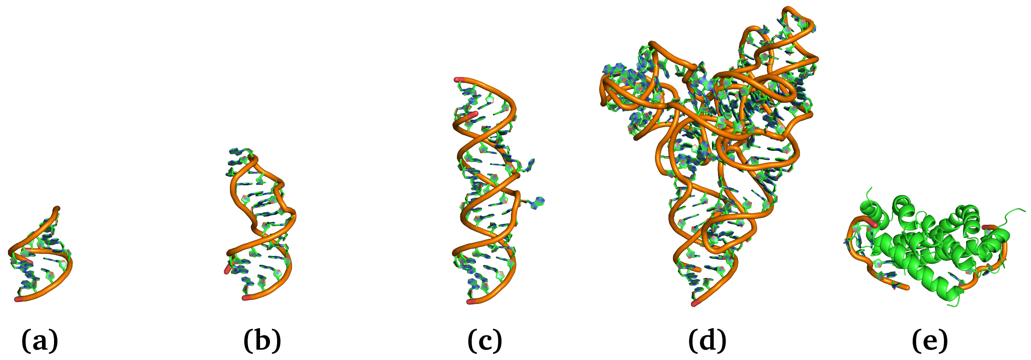


Figure II.1.1: The schematic display of the 4 RNA targets and the protein-DNA complex. (a) octameric RNA duplex, (b) sarcin ricin loop RNA, (c) Dimerization Initiation Site RNA, (d) GIR1 branching ribozyme, (e) σ^{E2}–TGTCAAA complex. The phosphate backbones are displayed in orange.

Table II.1.1: Number of nucleotides, molecular weight and motifs of the macromolecule targets

RNA	octameric RNA duplex (2 × 8-mer)	sarcin ricin loop (SRL)	Dimerization Initiation Site (DIS)	GIR1 branching ribozyme (GIR1)	σ^{E2}–TGTCAAA complex
Number of nucleotide	2 × 8	27	2 × 23	188	2 × 7
Molecular Weigth (kDa)	5.4	8.7	15.0	60.6	28.8
Motifs included	double helices	internal loop (bulged G)	double helices, internal loop	kissing stem-loop	protein-DNA complex

II.1.2 RNA synthesis, purification and crystallization

UCGUACGA octameric RNA duplex (2 × 8-mer) The RNA oligomers were ordered from Microsynth, Switzerland. The purification and crystallization were done by Joachim Schnabl, Institute of Inorganic Chemistry, Universität Zürich, Switzerland. The RNA was dissolved in 750 µl ddH$_2$O and mixed with 750 µl buffer. The purification was done with 200 µl 20% denaturing polyacrylamide gel. Prior to crystallization, the RNA of 1.2 µM was mixed 1:1 with a solution of 2.5 µM Co$_2$L$_3$ (L = C$_{25}$H$_{20}$N$_4$). Drops of 800 nl mixture were incubated at 293 K. Needle-shaped crystals grew after two days.

Sarcin ricin loop (SRL) The RNA were purchased from Dharmacon, Inc. (Boulder, CO, USA), purified and crystallized by Dr. Eric Ennifar, Institut de Biologie Moléculaire et Cellulaire, Université de Strasbourg, France, with protocols described previously.[90,91] Briefly, the RNA were purified by gel electrophoresis and concentrated to 2.5 µg µl^{-1} in 1.0 µM Na EDTA (pH 8.0), 10 µM Tris (pH 8.0). The solution was heated at 328 K and slowly cooled down to 298 K. The crystals were grown at 292 K in hanging drop by vapor diffusion with RNA/precipitant ratio 2:1. The well solution contained 3.0 to 3.2 M (NH$_4$)$_2$SO$_4$ mixed with buffer X (50 mM K · Mops (pH 7.0), 10 mM MgCl$_2$, 10 mM MnCl$_2$). Glycerol was used for cryoprotection. Crystals were frozen in ethane liquid.

Dimerization Initiation Site (DIS) Dimerization Initiation Site RNA (DIS) were purchased from Dharmacon, Inc. Purification and crystallization were done by Dr. Eric Ennifar, Institut de Biologie Moléculaire et Cellulaire, Université de Strasbourg, France. The purification and crystallization protocol have been described previously[92] and briefly recalled here. The RNA were purified by gel electrophoresis with 15% polyacrylamide. A further step purification was performed with a Dionex Nucleopac PA-100 ion-exchange column heated at 358 K. 10 times RNA solution containing 150 mM KCl , 20 mM Na cacodylate (pH 7.0), 5 mM MgCl$_2$

and RNA concentration of 4 μg μl^{-1} were mixed 1:1 with solution containing 10% MPD and 50 mM spermine. The reservoir solution contained 50% MPD, 300 mM KCl, 50 mM Na cacodylate (pH 7.0) and 100 mM $MgCl_2$. The crystals appeared after one week of setting the drop.

GIR1 branching ribozyme (GIR1) The RNA synthesis, purification and crystallization work was done by Mélanie Meyer, Institut de Biologie Moléculaire et Cellulaire, Université de Strasbourg, France.[93] The RNA were prepared by *in vitro* transcription using T7-RNA polymerase and concentrated by ethanol precipitation. A 100 μM RNA solution was heated during 1 min at 95 °C followed by 4 min on ice. $MgCl_2$ up to 4 μM was then added, the solution was placed at 55 °C for 10 min then at 4 °C overnight. Parallelepiped shaped crystals grew in 20 days in microbatch settings in 0.2 M sodium chloride, 0.1 M HEPES (pH 7.5), 25% *w/v* PEG 3,350. Crystals were flash-frozen by immersion in liquid ethane.

σ^{E2}–TGTCAAA complex Crystals of the σ^{E2}–TGTCAAA complex appeared spontaneously in the NMR tube with 10 μM Na-phosphate buffer pH 6.5, 50 μM NaCl at concentration of 0.5 μM.[94]

Chapter II.2

Experimental Set-up

II.2.1 X-ray source (X06DA/PXIII)

Diffraction data were collected at Beamline X06DA/PXIII at Swiss Light Source, Paul Scherrer Institut (SLS, PSI). The beamline received light from a 2.9 T super-bend magnet. The beam spot size at the source was 90 × 45 μm^2. The flux at 12.4 keV is 5×10^{11} phs/s which is comparable to an undulator beamline. Detailed source characteristics are shown in Table II.2.1.

Table II.2.1: Characteristics of the PXIII beamline at Swiss Light Source of Paul Scherrer Institut

Type	SLS super-bending magnet 2.9 Tesla
Source size (μm^2)	90×45
Spectral range (keV)	6.0 - 17.5
Flux (phs/s /400 μA)	5×10^{11}
Energy resolution (keV)	1.4×10^{-4}
Focused spot size (μm^2)	80×45

The beamline optics (Figure II.2.1) are designed to optimize stability. The beam is focused at the sample position 21 meters from the source. The vertical collimating mirror (M1) is focused at infinity and mounted at 7 meters from the source. Following the M1 mirror, a Double Channel Cut Monochromator (DCCM) allows a true fixed exit and thus enhances beam stability. The wide spectral range (6.0 to 17.5 keV) provides abundant choices of wavelength for experimental phasing. At 14 meters away from the source, a toroidal mirror (M2) suiting provides focusing in both vertical and horizontal directions with 2:1 horizontal demagnification.

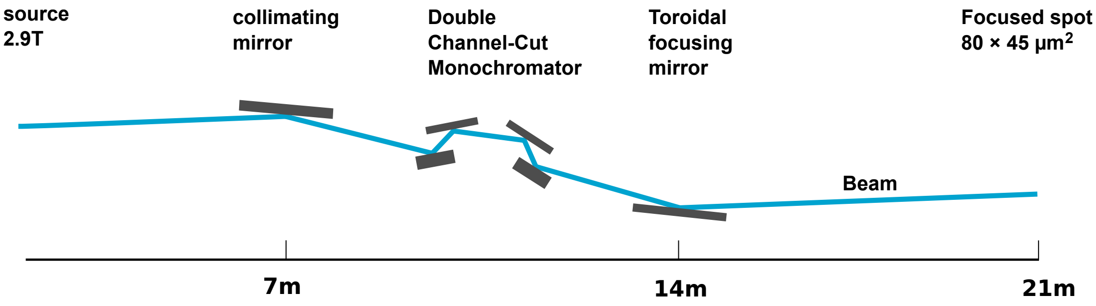


Figure II.2.1: The PXIII beamline layout

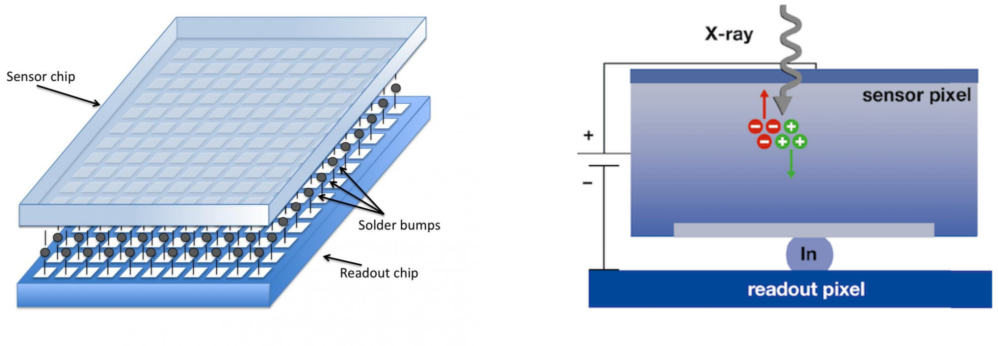
II.2.2 Detector (PILATUS 2M)

The detector equipped at beamline X06DA is a PILATUS-2M detector from DECTRIS, Ltd. PILATUS detectors operate in single-photon counting mode and feature very high dynamic range, zero dark signal and zero readout noise. With PILATUS detectors one can achieve optimal signal-to-noise ratio at short readout time and high frame rates.

The detector surface is made by grids of hybrid pixels. Electron charges are excited by the arriving photons collected on each pixel. The resulting current pulse is processed and amplified by the readout pixel in single-photon-counting mode (Figure II.2.2b).[95,96] Since the signal is recorded in digital only when it surpasses the defined limit, the noise raised by the dark current can be completely eliminated. In addition, the technique of on-fly photon counting while exposuring allows a fast digital readout and prevents readout noise which is a main source of errors in measuring with integrating detectors (Table II.2.2).[96–98]

---

[95] H. Spieler. *Semiconductor Detector Systems*. Oxford scholarship online OUP Oxford, 2005.
[96] DECTRIS *HYBRID PIXEL TECHNOLOGY & SINGLE-PHOTON COUNTING* 2012
[97] C. Ponchut. *J. Synchrotron Radiat.*, **13**: 195–203, 2006.
[98] P. Kraft. *PILATUS 2M*. PhD thesis. ETH Zürich, 2010.



(a) **(b)**

Figure II.2.2: The schematic of hybrid detector and single photon counting mechanism (a) Schematic layout of a hybrid detector (b) Principle of detection of X-ray photons counting in a hybrid pixel Detector

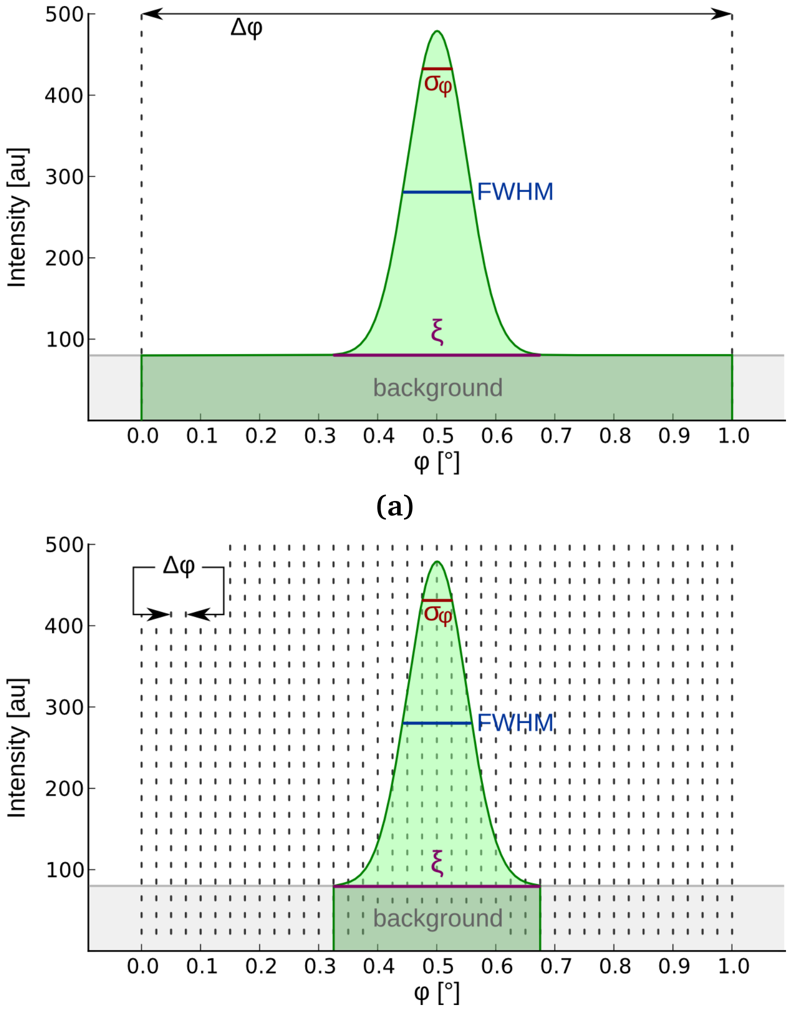
Table II.2.2: Detector characteristics of PILATUS-2M in comparison with the fiber optic coupled CCD detector

Characteristics	PILATUS 2M	CCD
Pixel size (μm^2)	172 × 172	46 × 46
Number of pixel	1475 × 1679	2048 × 2048
Dynamic range (bit)	20	14
Readout time (μs)	3.77	250
Frame rate (Hz)	60	4
Readout noise (e^-/pixel)	0	25.2
Dark noise (e^-/pixel)	0	0.18

Featuring zero readout noise, PILATUS detectors are not affected by dark current, which ensures a very accurate data collection. Low dose data collection is thus possible with PILATUS in order to minimize radiation damage. It is specifically important in high multiplicity data collection as the dose increases with multiplicity. The absence of readout noise together with the short readout time allows PILATUS detectors to collect data continuously and enables fine φ-slicing data collection[99] which may considerably improve the data quality. With fine φ-slicing data collection, the diffraction spots are able to be finer sampled and the optimal oscillation range were found to be around half of the mosaicity (Figure II.2.3).[99,100] Moreover, higher dynamic range of PILATUS compared to CCD detector (Table II.2.2) avoids the overload for data collected with high flux (Figure II.2.4).[96]

---

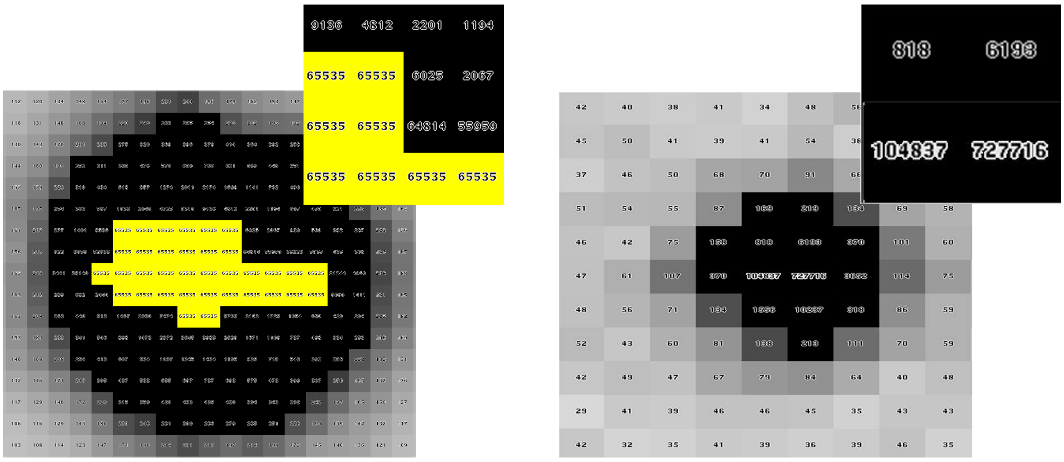
[99] M. Mueller et al. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **68**: 42–56, 2012.
[100] O. B. Zeldin et al. *J. Synchrotron Radiat.*, **20**: 49–57, 2013.



(a)

(b)

Figure II.2.3: Wide-sliced and fine-sliced data collection. The background and the reflection intensity along φ are shown, assuming a Gaussian distribution of the reflection intensity. (a) Wide-sliced data collection with a rotation width of 1°. The intensity of a full reflection (green outline) is recorded on a single image without sampling of the profile along φ. A large amount of background overlaps with the reflection intensity are included in the integration. (b) Fine-slicing at $\varphi = 0.025°$. The reflection profile is densely sampled along φ. The background not included in the reflection profile is reduced.[99]

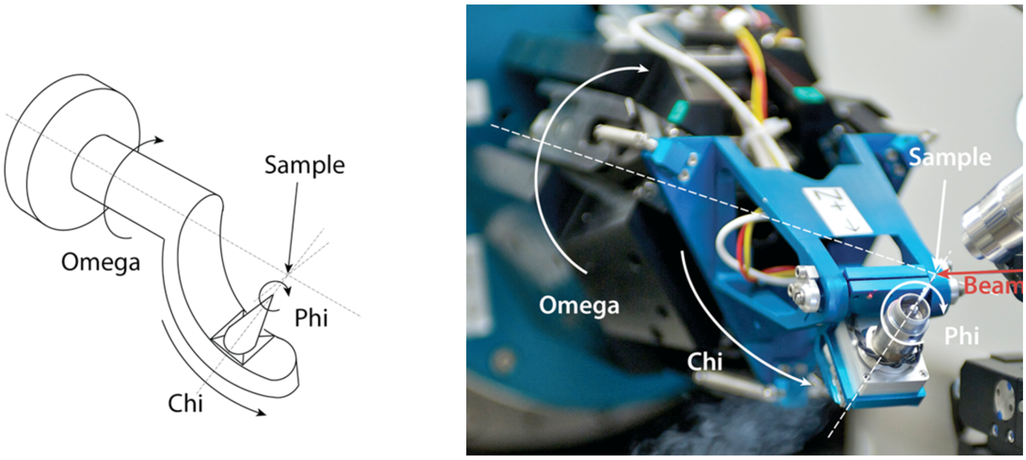


(a) **(b)**

Figure II.2.4: The diffraction spot measured on CCD and PILATUS detectors. (a) Diffraction spot collected on CCD detector with overloaded signal higher then the limit Analog/Digital Unit. (b) Diffraction spot recorded on PILATUS detector and the photon counting for each pixel. With higher dynamic range, there is no overload observed with the PILATUS detector.[96]

II.2.3 Multi-axes Goniometer PRIGo

The Parallel Robotics Inspired Goniometer (PRIGo) (Figure II.2.5) is a new goniometer concept, which offers an alternative for the traditional Kappa goniometer. It contains 6 degrees of freedoms including three translations (x, y, z) and three rotations (ω, φ, χ). The ω rotation is controlled by air-bearing rotation (ABR) and is followed by four linear stages to perform translations in three dimensions with χ rotation. The φ rotation is around the axis of the sample holder (Figure II.2.5). All motors are controlled synchronously by an elaborate inverse kinematics model implemented in the control system. The goniometer is designed to achieve the best accuracy while keep its flexibility to reach maximum orientation. The spheres of confusion for ω, χ and φ are < 5, < 7 and < 10 μm respectively. The rotation range for χ angle is 0-90° and for φ it is 0-360°. Unlike mini-Kappa goniometer which has a fixed angle between rotation axes, each axis of PRIGo can be tilted independently with higher flexibility.



(a) **(b)**

Figure II.2.5: Schematic architecture of PRIGo goniometer. (a) PRIGo emulating an arc with three rotation axes (ω, φ, χ) (b) PRIGo installed at beamline PXIII at Swiss Light Source

With large flexibility of crystal orientation, PRIGo allows the highest data completeness during data collection. For rotation data collection, not all possible diffraction can be measured, because some reciprocal lattice points lying close to the rotation axis will never cross the Ewald sphere, even after 360° rotation (Figure II.2.6).[101] The traditional way to bypass this restriction for complete data is to use bent loops. However, it is a crude approach without accurate control of the bending angle. With PRIGo, crystals can be orientated freely and accurately to bring the blind zone in crossing to the Ewald sphere and obtain high completeness data. Multiple choices of rotation axis provided by PRIGo enable multi-orientation data collection. High multiplicity data can be therefore collected around different rotation axes. Thus, systematic errors in data collection can be minimized. Furthermore, with the large range of χ and φ rotation, PRIGo allows collecting the *Bijvoet* pairs in a same image (*i.e.*, with the same X-ray dose) in order to measure more accurate anomalous differences.

---

[101] Z. Dauter. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **55**: 1703–1717, 1999.

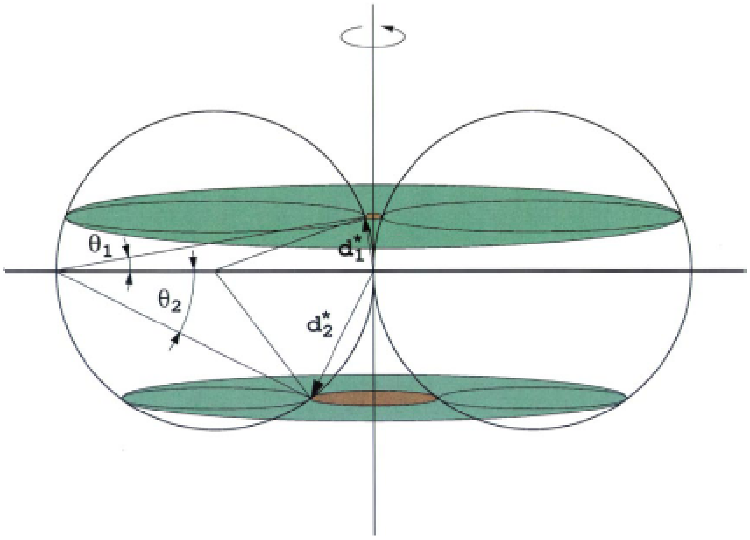


Figure II.2.6: A full 360° rotation of a crystal represented as the equivalent rotation of the Ewald sphere with the crystal stationary. Reflections in the blind region, close to the rotation axis, will never cross the surface of the sphere. Diffraction angles of θ_1 and θ_2 correspond to low and high resolution diffraction respectively. The smaller reciprocal interplanar spacing $d^*{}_1$ indicates lower resolution diffraction and the larger reciprocal interplanar spacing $d^*{}_2$ indicates higher resolution diffraction. The blind region is narrow at low resolution and wide at high resolution as shown by the red region in the figure. The green region indicates the measurable range of the diffraction at these two resolutions.[101]

Chapter II.3

Structure Determination

Structure determination can be divided in three consecutive steps: data-collection, phasing and refinement. X-ray data collection determines the quality of the crystallography data and therefore defines the success of the following procedures in the structure determination. The data collection strategies benefited from high brightness X-ray source, hybrid photon counting detector and muliti-orientation goniometer and used in each macromolecule target are given in this chapter. Phasing methods of SAD and MR and the refinement approach used in this thesis will also be described.

II.3.1 Data collection and processing

Diffraction data were collected at beamline X06DA at Swiss Light Source of Paul Scherrer Institut (section II.2.1) with samples described in section II.1.1. The crystals were measured under 100 K nitrogen-gas stream on a PILATUS 2M detector. With a 6 degrees of freedom (x,y,z,φ,ω,χ) multi-axis goniometer (section II.2.3), a high multiplicity and multi-axis data collection strategy was applied together with fine φ-slicing[99] to ensure true redundancy and optimize the data quality. Detailed data statistics is listed in Table II.3.1.

---

[99] M. Mueller et al. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **68**: 42–56, 2012.

2 × 8-mer RNA The crystal size is approximately 550 × 40 μm^2. Data were collected at 1.6 Å. 5 sub-datasets of 360° rotation were collected at different parts along the long needle like crystal. Each dataset were collected in fine φ-slicing along ω axis with χ tilting to different angles (χ = 0°, 10°, 20°, 30° and 40°). Limited by the detector size and sample-to-detector distance (120 mm), 2.11 Å data were collected. The crystal diffracts to higher resolution (1.2 Å).

SRL RNA The crystal size was around 400 × 300 μm^2. The wavelength was 1.6 Å. Highly redundant data were collected along ω with 0.1° oscillation fine φ-slicing.

DIS RNA Data were collected at 1.8 Å. Totally 5 sub-datasets of 360° rotation were collected at different parts of a same crystal. Each sub-dataset were collected along ω at 0.25° oscillation and 0.25 s exposure with χ axis tilted to different angles (χ = 0°, 10°, 20°, 30° and 40°).

GIR1 RNA The data were collected with 1.6 Å wavelength beam. Crystal size was around 160 × 25 μm^2. Totally 4 datasets of 360° rotation were collected with a same crystal. For each dataset, a fresh part of the crystal was shot. Two out of four sub-datasets were collected along ω. The other two sub-dataset were collected along φ axis with χ of 45° and 90° respectively.

σ^{E2}–TGTCAAA complex At wavelength of 2.0 Å, 720° data were collected with 0.1° oscillation and 0.1 s exposure. The total dose for this dataset was estimated to be about 1 MGy. Limited by the minimum sample-to-detector distance, useful data up to 2.14 Å were collected.

For data collected on each target (section II.1.1), *XDS* package was used to index the diffraction spots and integrate the intensities. The space group was then identified by *XDS*. *XSCALE* was used to scale and merge the multiple datasets. The merged data were then converted to appropriate format with *XDSCONV* in order to perform sub-structure determination and phasing. The absorption dose was estimated with *RADDOSE 3D*.[102]

---

[102] O. B. Zeldin et al. *J. Appl. Crystallogr.*, **46**: 1225–1230, 2013.

Table II.3.1: Diffraction data statistics for 5 macromolecule targets

Crystals	2 × 8-mer (2 × 7 P)	SRL (26 P)	DIS (2×22 P)	GIR1 (186 P)	σ^{E2}–TGTCAAA (12 P)
Wavelength (Å)	1.6	1.6	1.8	1.6	2.0
Resolution (Å)	1.7 (1.83-1.7)†	1.6 (1.8-1.6)†	2.1 (2.2-2.1)†	2.4 (2.5-2.4)†	2.2 (2.3-2.2)†
Exposure time (s)	0.2	0.1	0.25	0.25	0.1
Oscillation range (°)	0.2	0.1	0.25	0.25	0.1
χ angle (°)	0,10,20,30,40	0	0,10,20,30,40	0,0,90,45	0
Total Range (°)	1800	720	1800	1440	720
Number of images	9000	7200	7200	5760	7200
Redundancies	14.5	12.7	44.2	35.7	6.5
No. of Reflections					
Measured	138789	185096	642285	1439791	104807
Unique	9545	14566	14526	40378	16150
Dose per 360° rotation (MGy)	0.54	0.17	0.48	0.64	1.0
Space Group	$H3$	$P4_3$	$P3_12_1$	$P2_12_12_1$	$P2_1$
Unit-cell parameters					
a (Å), b (Å), c (Å)	46.90, 46.90, 53.16	29.61, 29.61, 76.08	59.05, 59.05, 64.28	57.63, 85.71, 108.53	39.01, 36.43, 73.27
α (°), β (°), γ (°)	90, 90, 120	90, 90, 90	90, 90, 90	90, 90, 90	90, 92.48, 90
Completeness (%)	95.2 (37.7)†	96.6 (79.9)†	99.8 (99.7)†	99.5 (98.7)†	72.6 (9.0)†
$\langle I/\sigma(I)\rangle$	57.5 (11.1)†	47.4 (22.8)†	58.8 (4.4)†	30.3 (1.5)†	60.0 (18.8)†
R_{meas} (%)‡	3.6 (11.0)†	4.0 (4.5)†	4.4 (82.4)†	12.3 (209.5)†	2.3 (4.6)†
$CC_{1/2}$ (%)	100 (98.1)†	100 (99.6)†	100 (87.4)†	99.9 (48.5)†	100 (99.7)†
CC_{ano} (%)	83	73	54	15	50

† The highest resolution shell

‡ $R_{\text{meas}} = \sum_{\mathbf{h}} \left(\frac{n_h}{n_h-1}\right) \sum_{l} |I_{hl} - \langle I_{\mathbf{h}}\rangle| / \sum_{\mathbf{h}} \sum_{l} \langle I_{\mathbf{h}}\rangle$

II.3.2 SAD phasing approach

Structure determination with SAD phasing approach includes two steps. First, the sub-structure composed of anomalous scatters needs to be determined. Then, with the known sub-structure, the phases of the whole molecule can be retrieved. Finally, the correct enantiomer can be chosen by assessing which hand for the anomalous scatters generates the better electron density map.[103] The detailed protocols used in phosphorus sub-structure determination and SAD phasing are described below.

II.3.2.1 Sub-structure determination

Package *SHELX2013*[104] was used to determine the sub-structures of the 5 targets. Before feeding into the *SHELX2013* package, merged data were first converted to HKLF4 format[105] with software *XDSCONV* in the *XDS* package.[106] Anomalous data to resolution limit were then prepared by *SHELXC* for running *SHELXD*. The method used to solve the sub-structure was combined Patterson searching and *Shake-and-Bake* dual-space refinement algorithm provided by *SHELXD*.[52] Depends on the specific tests, 1000 or 10000 search trials were attempted to solve phosphorus sub-structures. The best solution out of *SHELXD* was printed out as a PDB file[107] containing the coordinates of the putative substructure which can be used in SAD phasing.

The figure of merit used in *SHELXD* to judge the quality of a trial structure was the correlation coefficient (equation D.5.4) between the observed and calculated normalized *Bijvoet* differences. In *SHELXD*, $\mathrm{CC}_{\mathrm{all}}$ was the correlation coefficient calculated with all reflections. $\mathrm{CC}_{\mathrm{weak}}$ was the correlation coefficient obtained

---

[103] Z. Dauter et al. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **58**: 494–506, 2002.
[104] G. M. Sheldrick. *Acta Crystallogr. Sect. A Found. Crystallogr.*, **64**: 112–122, 2008.
[105] G. M. Sheldrick. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **66**: 479–485, 2010.
[106] W. Kabsch. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **66**: 125–132, 2010.
[52] T. R. Schneider and G. M. Sheldrick. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **58**: 1772–1779, 2002.
[107] PDB *Protein Data Bank Contents Guide* tech. rep. 2008, 205

with reflections not using in localization of the anomalous scatters. It was a concept analogous to R_{free} and was used to monitor over-fitting and control the data validation. The proportion of free data in CC_{weak} calculation for SAD/MAD sub-structure determination in *SHELXD* should be at least 15%.[108] The successful solution typically showed high values for both CC_{all} and CC_{weak}. Random solutions may have high CC_{all} but very low CC_{weak} ($< 10\%$).[109] A typically sign for the successful *SHELXD* sub-structure localization is the separation of clusters in the CC_{all} vs. CC_{weak} plot. The solution of *SHELXD* was selected and ranked based on their CC_{all} and CC_{weak} values.

To evaluate the solution explicitly, *SITCOM*[110] were used to directly compare *SHELXD* solutions with the known sub-structure coordinates. A list of unique sub-structure sites were constructed in *SITCOM* based on the known phosphorus sub-structure. Then the target solution was mapped to the unique sites by symmetry operation or origin shifts. While the fitting was done, a site figure of merit (SFOM) was assigned to the identified sites (Appendix D.6). The correct sites should have SFOM greater than 50% and d_{mean} smaller than 1.0 Å.

II.3.2.2 Phasing and density modification

Initial SAD phases were obtained by *PHASER*[89] SAD pipeline. In *PHASER*, determined sub-structures and structure factors of *Bijvoet* pairs were given as input. The structure sequences were provided to estimate the solvent content.

Phase ambiguities of the outputs from *PHASER* were resolved by performing density modification with *DM*.[111,112] Number of cycles of phase extension in *DM* was set to AUTO mode in which the phase extension in *DM* terminated until the real-space-free residual stopped decreasing. This setup avoided over-weighting

---

[108] T. Grüne *Experimental Phasing with shelx c/d/e - Phasing in the Context of Structure Determination* tech. rep. 2013

[109] G. M. Sheldrick. , , 2007.

[110] F. Dall'Antonia and T. R. Schneider. *J. Appl. Crystallogr.*, **39**: 618–619, 2006.

[89] A. J. McCoy et al. *J. Appl. Crystallogr.*, **40**: 658–674, 2007.

[111] K. Cowtan. *Jt. CCP4 ESF-EACBM Newsl. Protein Crystallogr.*, **31**: 34–38, 1994.

[112] K. Cowtan and K. Y. J. Zhang. *Prog. Biophys. Mol. Biol.*, **72**: 245–270, 1999.

and bias caused by calculation for too many cycles.

The Electron density maps out of density modification were then evaluated in the visualization software *COOT*.[113] To evaluate the phase quality, map correlation coefficient of the experimental phasing map against the map generated from refined models (map CC) were calculated with *phenix.model_vs_map*.[114,115]

---

[113] P. Emsley and K. Cowtan. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **60**: 2126–2132, 2004.
[114] P. D. Adams et al. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **66**: 213–221, 2010.
[115] P. V. Afonine et al. *J. Appl. Crystallogr.*, **43**: 669–676, 2010.

II.3.3 Molecular Replacement

In Molecular Replacement (MR), homologous models are used to search the correct solution via rotation and translation in unit-cell. While the initial solution is obtained, iterative refinement can be performed to improve the quality of the model. The MR approaches used in this thesis and the concept of *de novo* MR search using model fragments are described in the following sections. The strategies of using anomalous signals in combination with MR to facilitate structure determination will also be reviewed.

II.3.3.1 MR search and solution evaluation

PHASER[89] was used to perform Molecular Replacement search. The sequence files were used to estimate the solvent content in *PHASER*. To evaluate the quality of the solution, translation function Z-score (TFZ) calculated with the top solution after refinement were checked. Based on the previous experiences of MR practices with *PHASER*, rough quantitative guide to interpreting TFZ score is listed in Table II.3.2.[116] To further evaluate the solutions, maps and models out of *PHASER* were checked in *COOT*[113] and the solutions out of *PHASER* were compared to the true model using a *Mathematica*[117] script based on the *BioCrystallographica* package.[118]

Table II.3.2: The guideline for interpreting the TFZ scores of the MR solutions from *PHASER*

TFZ score	Solved or Not
< 5	no
5 - 6	unlikely
6 - 7	possibly
7 - 8	probably
> 8	definitely

---

[116] A. J. McCoy. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **63**: 32–41, 2007.

[117] Wolfram Research Inc. *Mathematica Edition: Version 9.0* Champaign, Illinois, 2012

[118] N. Ambert et al. *CCP4 Newsl.*, **44**: 22–31, 2006.

II.3.3.2 Concept of *de novo* MR

It was shown that RNA model fragments can be used in MR with iterative cycles of model correction and refinement to determine the RNA structure.[2,119] Similar concept of applying *de novo* structure determination with model fragments is showed in Figure II.3.1. First, the secondary structure of the target RNA is required. It can be predicted with the software *ASSEMBLE*.[120] Then 3 dimensional model fragments can be built based on RNA motifs in the secondary structural. Since abundant sequences are nowadays available in databases, a fragments library can be built through *ASSEMBLE* server with source data in the Protein Data Bank.[13] Then the fragments in the library can be used in independent or ensemble MR search. If the partial solution were obtained from *PHASER* as indicated by significant TFZ scores, iterative MR search with subsequent model corrections and refinement can be applied to extend the phases for solving the whole structure.

---

[2] M. P. Robertson and W. G. Scott. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **D64**: 738–744, 2008.

[119] M. P. Robertson et al. *Methods*, **52**: 168–172, 2010.

[120] F. Jossinet et al. *Bioinformatics*, **26**: 2057–2059, 2010.

[13] H. M. Berman et al. *Nucleic Acids Res.*, **28**: 235–242, 2000.

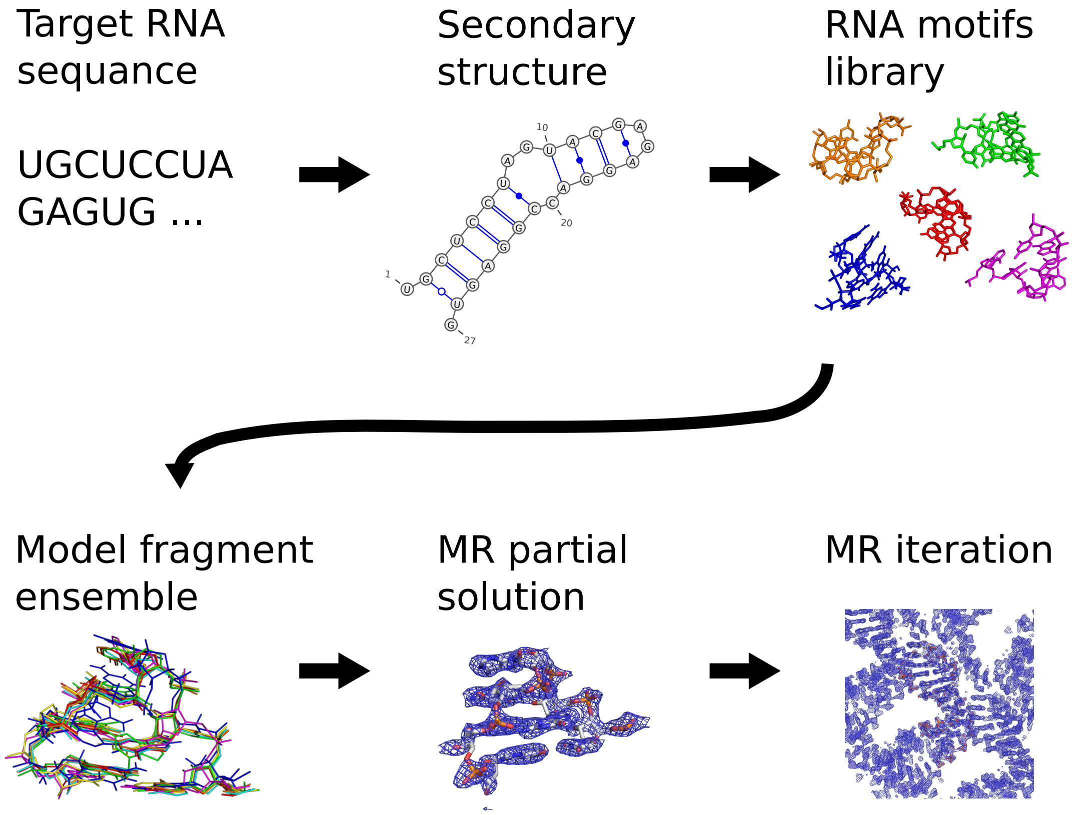


Figure II.3.1: The strategy of applying *de novo* MR phasing for the nucleic acid structures structure determination

II.3.3.3 Using anomalous information in combination with MR

The anomalous signals can be used together with search model in MR to enhance phasing power. One strategy is to use MR search with the anomalous model against anomalous differences $|\Delta_{anom}|$. The anomalous model is solely composed of the atoms of anomalous scatters. In the case of nucleic acid, this is the model constituted of all phosphorous atoms in the structure.

Another strategy was to use the partial model fragments to initiate MR-P-SAD search in *PHASER*. For MR-P-SAD in *PHASER*, SAD log-likely gradient map can be used to complete the sub-structure. The extended sub-structure can be then used iteratively with *PHASER* or *SHELXE* to phase the whole structure[2,119] (Figure II.3.2).

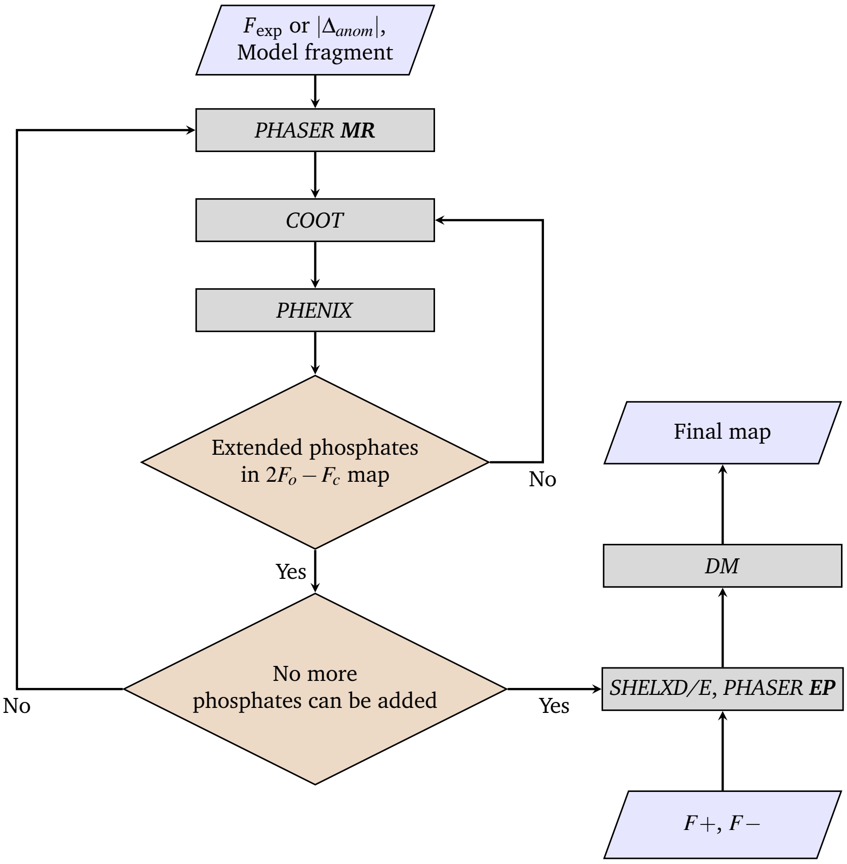


Figure II.3.2: Concept of the iterative Molecular Replacement (MR) used in combination with anomalous information for nucleic acid structure determination. To perform initial MR search, either experimental amplitudes (F_{exp}) or the anomalous differences ($|\Delta_{anom}|$) can be used. With the amplitudes, normal MR can be used to determine the orientation and position of the initial model fragment. With anomalous differences, the phosphorus substructure from the search model will be used in MR search. The initial solution of MR is then checked and corrected in *COOT* and refined in *PHENIX*. If additional information in the density map ($2F_o - F_c$) at the fresh parts is observed, the solution can be extended. Additional phosphates are then built on to the MR solution at the fresh part. The extended model with newly added phosphates can be used as a search model in another round of MR search. Iteration of this procedure continues until no more phosphates can be added. The phosphorus extracted from the final model can be then used as the initial heavy-atom seed for *SHELXD*. The refined sub-structure and measured amplitudes of Friedel pairs ($F+$, $F-$) can be then used in *SHELXE* or *PHASER* for the experimental phasing.

II.3.4 Refinement

The 8-mer and Protein-DNA complex model were built in electron density map obtained from density modification after experimental phasing. SRL and DIS models were obtained with the previously determined structures[92,121] in Protein Data Bank (PDB).[13] Refined structures of the 5 targets were used for the results evaluation of sub-structure determination, experimental phasing and Molecular Replacement. The refinement of SRL RNA, 8-mer RNA and Protein-DNA complex were performed with *phenix.refine*[122] in *PHENIX* package.[114] The refinement of GIR1 was done by Mélanie Meyer, Institut de Biologie Moléculaire et Cellulaire, Université de Strasbourg.[93] The refinement statistics are shown in Table II.3.3.

---

[92] E. Ennifar et al. *Structure*, **7**: 1439–1449, 1999.
[121] C. C. Correll et al. *Nucleic Acids Res.*, **31**: 6806–6818, 2003.
[122] P. V. Afonine et al. *CCP4 Newsl.*, **42**: 1–18, 2005.
[93] M. Meyer et al. *Proc. Natl. Acad. Sci. U. S. A.*, **111**: 7659–7664, 2014.

Table II.3.3: Refinement statistics

	2 × 8-mer	SRL	DIS	GIR1	σ^{E2}–TGTCAAA
Resolution range (Å)	32.27-1.66	23.37-1.60	40.02-2.10	47.82-2.40	38.97-2.14
R factor	0.189	0.140	0.193	0.233	0.164
R_{free}	0.196	0.172	0.246	0.281	0.217
No. of non-H atoms	371	803	1025	4225	1864
No. of solvent atoms	35	105	39	220	100
R.m.s.d from ideal geometry					
Bond length (Å)	0.004	0.004	0.007	0.006	0.008
Angle distances (Å)	0.904	0.922	1.189	1.325	0.982
Ramachandran plot (%)					
Favored	–	–	–	–	98.3
Allowed	–	–	–	–	1.7
Outliers	–	–	–	–	0

Part III

Results and Discussion

Chapter III.1

Possibilities and Limits of P-SAD

In order to study the feasibility of the P-SAD method, systematic analyses were performed on experimental and simulated data. The effect of wavelength, multiplicity on data quality, sub-structure determination and phasing will be discussed in the following chapter.

III.1.1 Effect of wavelength on P-SAD

III.1.1.1 Effect of wavelength on data quality

Atomic scattering factor contains wavelength dependent contributions $f'_{(\lambda)}$ and $f''_{(\lambda)}$. Therefore, anomalous signals can be enhanced with proper wavelength. As the phosphorus absorption K-edge (2143 eV) is not achievable at current beamline set-up, the only way to increase anomalous scattering is to use longer wavelength, *i.e.*, lower energy beam. However, since energy absorption increases radiation damage,[123] data quality may degrade at longer wavelength. In addition to wavelength (or energy), factors such as the crystal size, cryotechniques and other instrumental factors can influence the measurement of intensities and anomalous differences.[123] Here, anomalous signals of SRL data collected at different wavelength (1.2 to 2.2 Å) are presented to show the overall effect of wavelength on anomalous signals.

---

[123] I. Polikarpov et al. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **53**: 734–737, 1997.

Data statistics showed that data collected at different wavelength of similar quality (Table III.1.1). To evaluate the anomalous strength at each wavelength, the correlation coefficients between experimental and calculated anomalous differences were plotted against resolution (Figure III.1.1).

Increased correlations were observed at longer wavelength showing the benefit of collecting anomalous data at longer wavelength. However, according to Bragg's law, the diffraction angle is proportional to the wavelength.[a] Therefore, for data collected at long wavelength, the measurable range of resolution is limited by the detector distance/size/geometry. Due to this limitation, high resolution diffraction may not be measured at long wavelength. 1.6 Å wavelength was then chosen as a compromise between anomalous strength and loss of high resolution data.

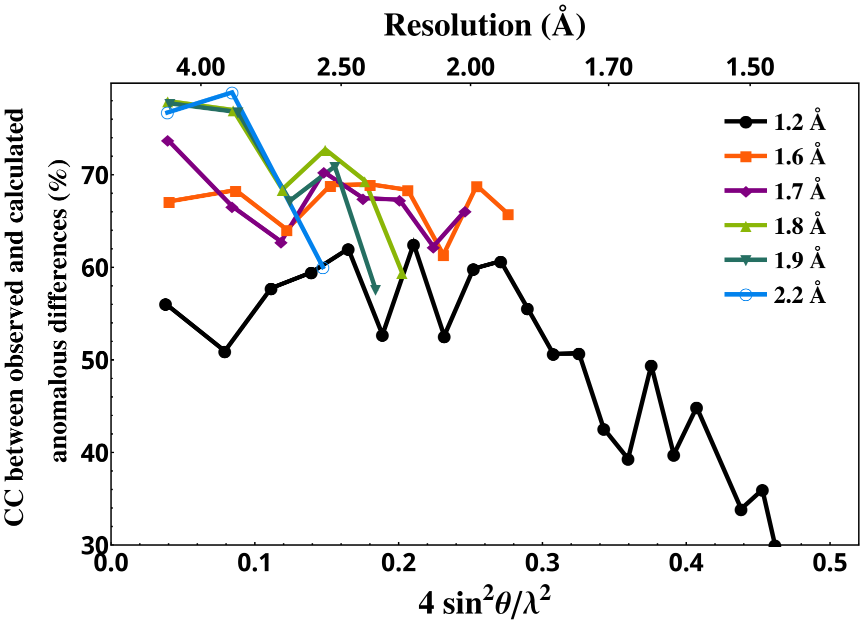


Figure III.1.1: Correlation coefficients between the experimental measured anomalous differences and that calculated from refined models at measurement wavelength from 1.2 to 2.2 Å

---

[a] $2d\sin\theta = n\lambda$, where θ is the scattering angle and λ is the wavelength. Thus angle of scattering from a crystal lattice with long wavelength is larger than short wavelength.

Table III.1.1: Crystallography data statistics for data collected at different wavelength

Wavelength (Å)	**1.2**	**1.6**	**1.7**	**1.8**	**1.9**	**2.2**
Resolution (Å)	1.2 (1.3-1.2)[†]	1.8 (1.9-1.8)[†]	2.0 (2.1-2.0)[†]	2.1 (2.2-2.1)[†]	2.25 (2.35-2.25)[†]	2.6 (2.7-2.6)[†]
Exposure(s)	0.5	1	1	1	1	0.2
Oscillation (°)	0.5	1	1	1	1	0.2
Rotation range (°)	720	720	720	720	720	720
Space Group	$P\,4_3$	$P\,4_3$	$P\,4_3$	$P\,4_3$	$P\,4_3$	$P\,4_3$
Unit-cell parameters						
a (Å)	29.55	29.61	29.59	29.47	29.48	29.52
b (Å)	29.55	29.61	29.59	29.47	29.48	29.52
c (Å)	76.36	76.12	76.06	75.85	75.97	76.22
α (°)	90	90	90	90	90	90
β (°)	90	90	90	90	90	90
γ (°)	90	90	90	90	90	90
No. of reflections						
Measured	589858	146540	126006	109548	88679	47220
Unique	41347	10155	8700	7439	6083	3977
Multiplicities	14.3	14.5	14.5	14.7	14.6	11.9
Completeness (%)	99.8 (100)[†]	99.9 (100)[†]	99.9 (100)[†]	100 (100)[†]	100 (100)[†]	100 (100)[†]
R_{merge}(%)[‡]	3.1 (19.3)[†]	3.3 (4.2)[†]	3.4 (4.9)[†]	2.9 (6.5)[†]	2.4 (6.2)[†]	3.2 (4.0)[†]
R_{meas}(%)[§]	3.2 (20.1)[†]	3.4 (4.3)[†]	3.6 (5.1)[†]	3.0 (6.8)[†]	2.5 (6.4)[†]	3.3 (4.2)[†]
$\langle I/\sigma(I)\rangle$	47.6 (14.0)[†]	68.6 (51.4)[†]	61.9 (42.6)[†]	73.9 (38.2)[†]	86.1 (42.4)[†]	65.7 (44.1)[†]
$\text{CC}_{1/2}$(%)	100 (99.2)[†]	100 (99.9)[†]	100 (99.9)[†]	100 (99.9)[†]	100 (99.9)[†]	100 (99.9)[†]
CC_{ano}(%)	31 (8)[†]	80 (70)[†]	77 (66)[†]	77 (57)[†]	83 (54)[†]	91 (88)[†]

[†] Highest resolution shell

[‡] $R_{\text{merge}} = \sum_{\mathbf{h}}\sum_{l}|I_{hl} - \langle I_{\mathbf{h}}\rangle| / \sum_{\mathbf{h}}\sum_{l}\langle I_{\mathbf{h}}\rangle$

[§] $R_{\text{meas}} = \sum_{\mathbf{h}}\left(\frac{n_h}{n_h-1}\right)\sum_{l}|I_{hl} - \langle I_{\mathbf{h}}\rangle| / \sum_{\mathbf{h}}\sum_{l}\langle I_{\mathbf{h}}\rangle$

III.1.1.2 Effect on sub-structure determination

Due to the experimental limitation, high resolution reflections cannot be measured at longer wavelength. The limitation in resolution can however affect sub-structure determination and phasing. To study the resolution limit for the phosphorus sub-structure determination, 10000 *SHELXD*[52,104] search trials were performed with SRL data at different resolution cut-off from 1.8 to 3.0 Å. To evaluate the *SHELXD* results at different resolution cut-off, CC_{all} (section II.3.2.1) were plotted against CC_{weak} (Figure III.1.2). For data with low resolution cut-off (2.0 - 3.0 Å), only random solutions were found. At resolution higher than 2.0 Å, correct solutions indicated by high CC_{all}/CC_{weak} were observed. The numbers of correct sites in *SHELXD* solution at different resolution cut-off are listed in Table III.1.2. It was found that resolution higher than 2.0 Å was required to solve the sub-structure.

Table III.1.2: The numbers of correct sites in *SHELXD* solutions for data cut at different resolutions

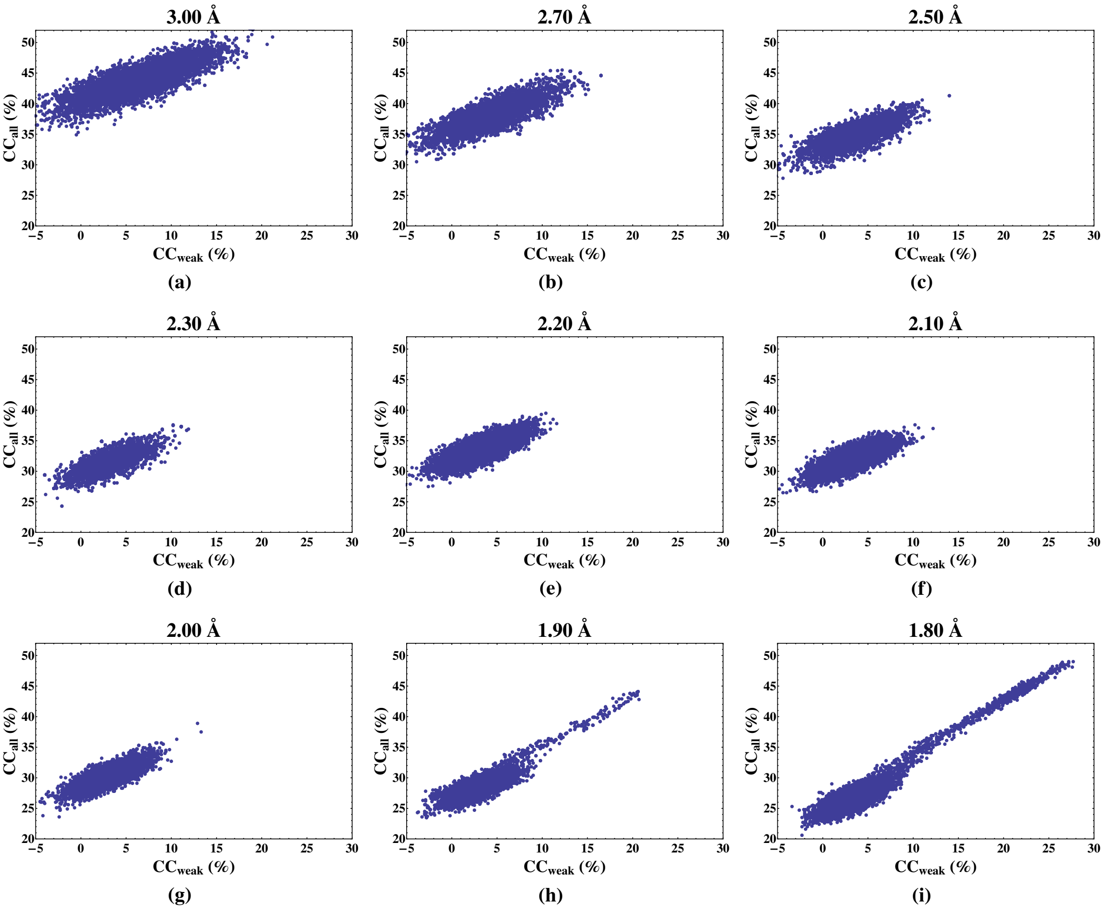
Resolution (Å)	1.80	1.90	2.00	2.10	2.20	2.30	2.50	2.70	3.00
Number of sites	26	23	18	4	3	1	3	1	2

To further analyze the effect of the resolution limit on sub-structure determination, experimental as well as simulated data for 4 RNA targets were cut at different resolutions and then used in *SHEXLD* for sub-structure determination. The result of sub-structure determination (success/fail) is listed in Table III.1.3. For all targets except GIR1 RNA data, higher than 2.8 Å data were required to determine the sub-structure. The sub-structure of GIR1 RNA containing 186 sites could not be solved at any resolution cut-off within 1.8 to 3.0 Å.

---

[52] T. R. Schneider and G. M. Sheldrick. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **58**: 1772–1779, 2002.

[104] G. M. Sheldrick. *Acta Crystallogr. Sect. A Found. Crystallogr.*, **64**: 112–122, 2008.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure III.1.2: *SHELXD* $\mathrm{CC_{all}}$ vs. $\mathrm{CC_{weak}}$ plots for sub-structure determination with SRL data cut at different resolutions

Table III.1.3: *SHELXD* sub-structure determination solutions with resolution cut-off from 1.8 to 2.5 Å for 4 RNA targets. Successfully determined sub-structures were marked as ✓ and the failed ones were marked as ✗.

(a) Experimental data

Molecule	Number of P	Resolution (Å)									
		1.8	2.0	2.1	2.2	2.3	2.4	2.5	2.7	2.8	3.0
2×8-mer	14	✓	✓	✓	✓	✓	✓	✗	✗	✗	✗
SRL	26	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗
DIS	44	–	–	✗	✗	✗	✗	✗	✗	✗	✗
GIR1	186	–	–	–	–	–	✗	✗	✗	✗	✗

(b) Simulated data

Molecule	Number of P	Resolution (Å)									
		1.8	2.0	2.1	2.2	2.3	2.4	2.5	2.7	2.8	3.0
2×8-mer	14	✓	✓	✓	✓	✓	✓	✓	✓	✗	✗
SRL	26	✓	✓	✓	✓	✗	✗	✗	✗	✗	✗
DIS	44	✓	✓	✓	✗	✗	✗	✗	✗	✗	✗
GIR1	186	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗

III.1.1.3 Effect on P-SAD phasing

The effect of resolution on P-SAD phasing was analyzed. Phases were determined with *PHASER*[89] with the known sub-structure and improved by density modification with *DM*[112] as described in section II.3.2.2. More accurate phases were obtained with higher resolution cut-off (Figure III.1.3). For data with resolution higher than 2.0 Å, less than 40° phase errors were observed after density modification. This phase difference corresponds to map CC $> 77\%$ (map CC is the $|F|^2$-weighted average of the values of $\cos(\Delta\varphi)$[124]). For data cut-at 2.5 Å, the maximum phase error is 50°which corresponds to map CC of 64%. Interpretable maps were obtained at resolution cut-off of 2.5 Å or higher (Figure III.1.4). We found that sub-structure determination of SRL experimental data requires at least 2.0 Å resolution (section III.1.1.2). Compared to sub-structure determination, we observed that phasing can be performed at lower resolution (2.5 Å).

---

[89] A. J. McCoy et al. *J. Appl. Crystallogr.*, **40**: 658–674, 2007.

[112] K. Cowtan and K. Y. J. Zhang. *Prog. Biophys. Mol. Biol.*, **72**: 245–270, 1999.

[124] V. Y. Lunin and M. M. Woolfson. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **49**: 530–533, 1993.

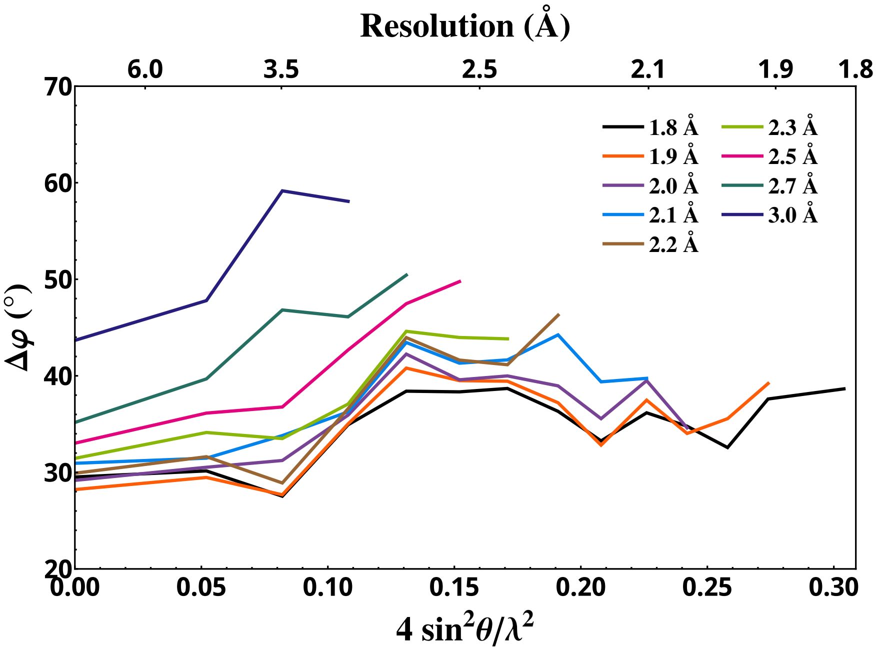


Figure III.1.3: Phase differences ($\Delta\varphi$ in degree) between the calculated phases and the phases out of experimental phasing after density modification. The experimental phasing was performed with SRL data at resolution cut-off from 1.8 to 3.0 Å in *PHASER*. Density modification was done in *DM*. The calculated phases were obtained from the final refinement model at 1.6 Å.

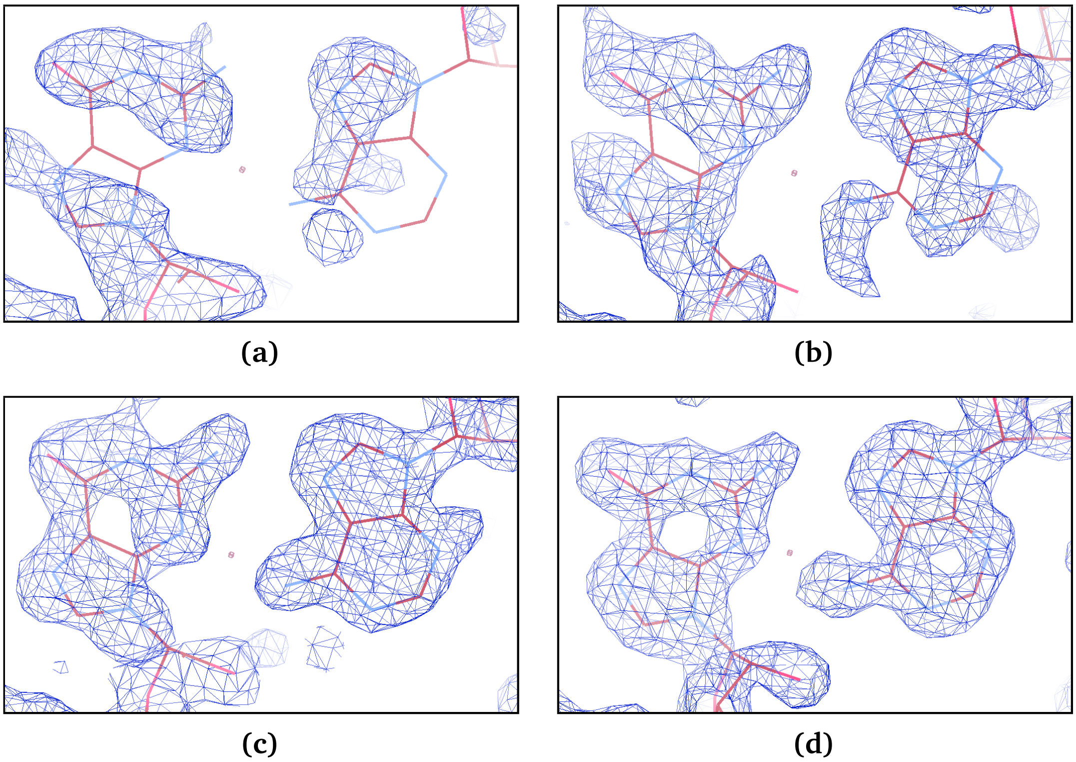


Figure III.1.4: Experimental density map of data cut at different resolution: (a) **3.0** Å, (b) **2.8** Å, (c) **2.5** Å, and (d) **1.8** Å. All maps were obtained after density modifications and contoured at 1.0 σ.

III.1.2 Effect of multiplicity on P-SAD

High multiplicity data is crucial for the success of P-SAD phasing.[1] In order to analyze the effect of multiplicity on data quality, phosphorus sub-structure determination and P-SAD phasing were performed systematically at different multiplicities with SRL (26 nt) data (section II.3.1). For this, SRL data were divided into eight successive wedges based on the oscillation range (Table III.1.4).

---

[1] Z. Dauter and D. A. Adamiak. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **57**: 990–995, 2001.

Table III.1.4: Data statistics for data splited at different multiplicities

Data set	**120**	**180**	**270**	**360**	**450**	**540**	**630**	**720**
Resolution (Å)	1.8 (1.9-1.8)†	1.8 (1.9-1.8)†	1.8 (1.9-1.8)†	1.8 (1.9-1.8)†	1.8 (1.9-1.8)†	1.8 (1.9-1.8)†	1.8 (1.9-1.8)†	1.8 (1.9-1.8)†
Rotation range (°)	120	180	270	360	450	540	630	720
No. of reflections								
Measured	24491	36924	55438	73966	92420	110578	128435	146540
Unique	9879	10002	10137	10133	10153	10153	10153	10155
Multiplicities	2.5	3.6	5.3	7.3	9.1	10.9	12.7	14.5
Completeness (%)	97.2 (98.2)†	98.4 (98.3)†	99.8 (100)†	99.8 (99.8)†	99.8 (100)†	99.8 (100)†	99.8 (100)†	99.9 (100)†
R_{merge} (%)‡	2.9 (3.8)†	3.1 (3.8)†	3.3 (4.2)†	3.3 (4.2)†	3.2 (4.1)†	3.2 (4.1)†	3.3 (4.1)†	3.3 (4.2)†
R_{meas} (%)§	3.7 (4.6)†	3.6 (4.5)†	3.6 (4.5)†	3.6 (4.4)†	3.4 (4.4)†	3.4 (4.3)†	3.4 (4.3)†	3.4 (4.3)†
$\langle I/\sigma(I)\rangle$	25.4 (18.8)†	32.3 (24.2)†	39.2 (29.6)†	45.9 (33.6)†	53.1 (38.8)†	59.2 (43.9)†	64.4 (48.1)†	68.6 (51.4)†
$CC_{1/2}$ (%)	99.8 (99.7)†	99.9 (99.8)†	99.9 (99.8) †	99.9 (99.9)†	100 (99.0)†	100 (99.9)†	100 (99.9)†	100 (99.9)†
CC_{ano} (%)	30 (25)†	49 (43)†	55 (46)†	62 (51)†	69 (58)†	74 (65)†	77 (68)†	80 (70)†

† Highest resolution shell

‡ $R_{\text{merge}} = \sum_{\mathbf{h}}\sum_{l}|I_{hl} - \langle I_{\mathbf{h}}\rangle| / \sum_{\mathbf{h}}\sum_{l}\langle I_{\mathbf{h}}\rangle$

§ $R_{\text{meas}} = \sum_{\mathbf{h}}\left(\frac{n_h}{n_h-1}\right)\sum_{l}|I_{hl} - \langle I_{\mathbf{h}}\rangle| / \sum_{\mathbf{h}}\sum_{l}\langle I_{\mathbf{h}}\rangle$

III.1.2.1 Effect of multiplicity on data quality

Statistical indicators of data qualities were checked at different multiplicities. Statistical measures for the accuracy in intensities are the R_{meas} and the intensity to noise ratio $I/\sigma(I)$. The indicators of the anomalous strength are the *Bijvoet* ratio $\Delta F_{\text{ano}}/F$ and the anomalous correlation coefficient CC_{ano}.

- **Effect of multiplicity on the measurement of intensities**

  The effect of multiplicity on the measurement of intensities was showed by comparison of R_{meas} and intensity to noise ratio of data at different multiplicities. R_{meas} (equation D.5.2) is a measure of consistency in dataset which is independent of data multiplicity.[125] Lower R_{meas} indicates better precision of intensities. R_{meas} for higher multiplicity data were found to be lower than that of lower multiplicity data showing the improvement of data accuracy at higher multiplicity (Figure III.1.5).

  As showed in the Figure III.1.6, data with higher multiplicity had higher signal to noise ratio (Figure III.1.6). The average $I/\sigma(I)$ value of data **720** (corresponding to 720° rotation) was approximately twice as high as that of data **180**. This is in accordance with the statistical expectation that the standard deviation decreased by a factor that proportional to the square-root of measurement multiplicity.[1]

---

[125] K. Diederichs and P. A. Karplus. *Nature*, **4**: 269–275, 1997.

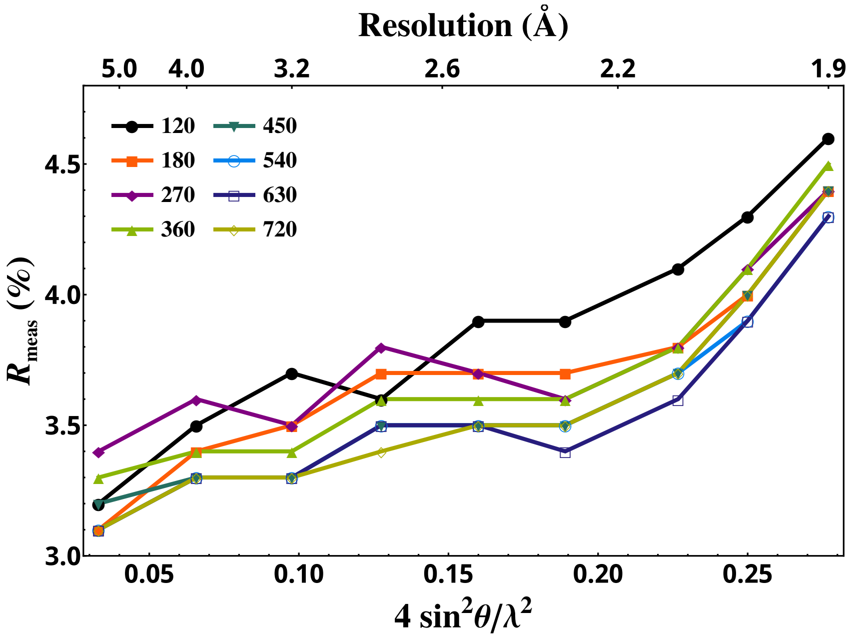


Figure III.1.5: R_{meas} as a function of resolutions for 8 wedges of data with different multiplicities

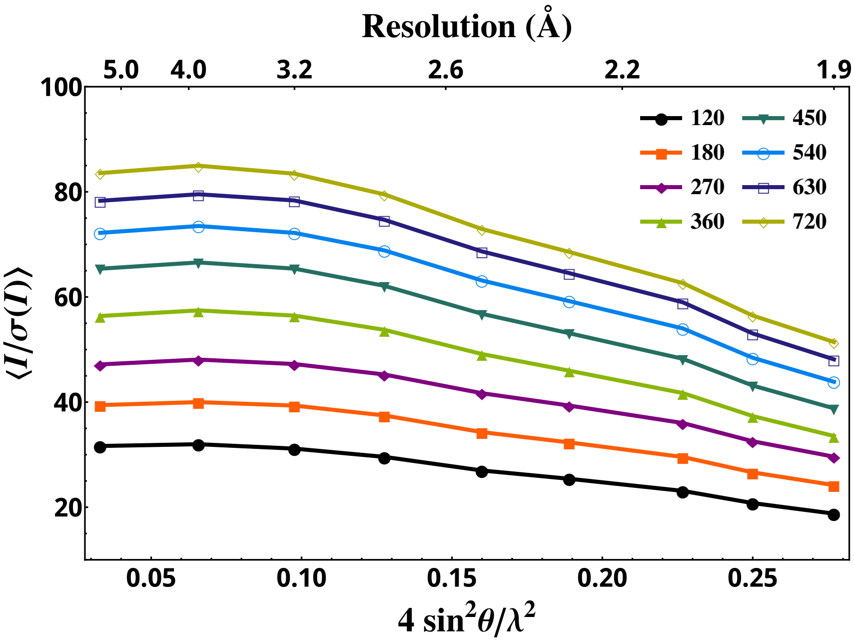


Figure III.1.6: $\langle I/\sigma(I)\rangle$ in respect to resolution for the 8 wedges of data with different multiplicities

- **Effect of multiplicity on the measurement of anomalous signals**

To access the influence of data multiplicity on the measurements of anomalous signals, the average *Bijvoet* ratio $\Delta F_{\text{ano}}/F$ was calculated for each wedge of the experimental data and plotted against resolution (Figure III.1.7).

Measurements errors cause $\Delta F_{\text{ano}}/F$ to tend towards unity, which does not indicate a higher amount of the anomalous strength. As seen in Figure III.1.7, the higher the multiplicity the closer the experimental curves to the theoretical estimation, which confirms the improvement in anomalous strength of higher multiplicity data.

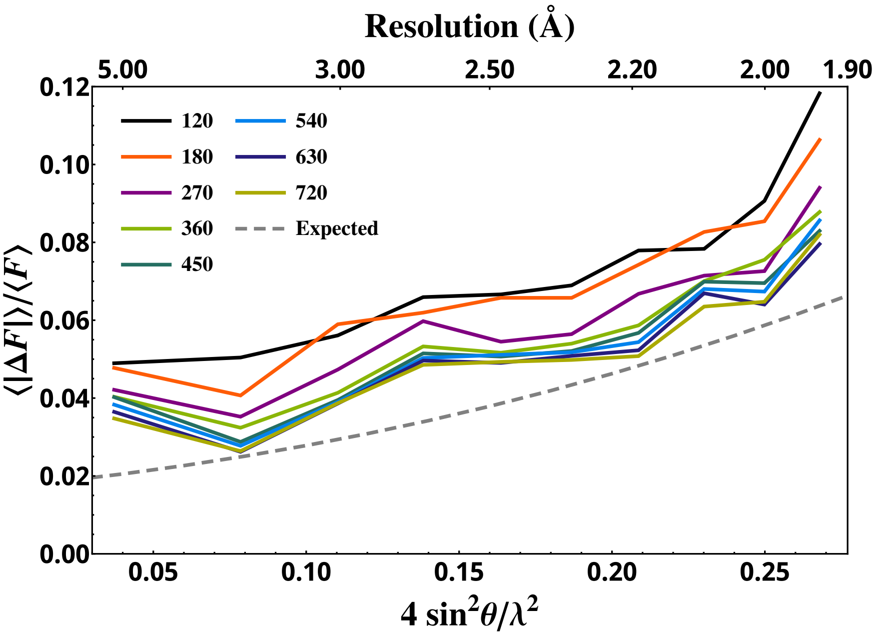


Figure III.1.7: $\Delta F_{\text{ano}}/F$ against resolution for data sets with different multiplicities; The estimated anomalous signals were calculated based on the known number of anomalous scatters and the total number of atoms (equation D.5.3) and then weighted by the corresponding temperature factors.

CC_{ano} is another parameter to evaluate the accuracy of anomalous measurements. The CC_{ano} versus resolution is shown in Figure III.1.8. The enhancement of the anomalous signal with multiplicity is showed by the increased CC_{ano}. For low multiplicity data (**120**), the overall CC_{ano} was lower than 40% and felt below 30% at high resolution range. High multiplicity data (**450** to **720**) had CC_{ano} value larger than 50% even at resolution range beyond 2.0 Å.

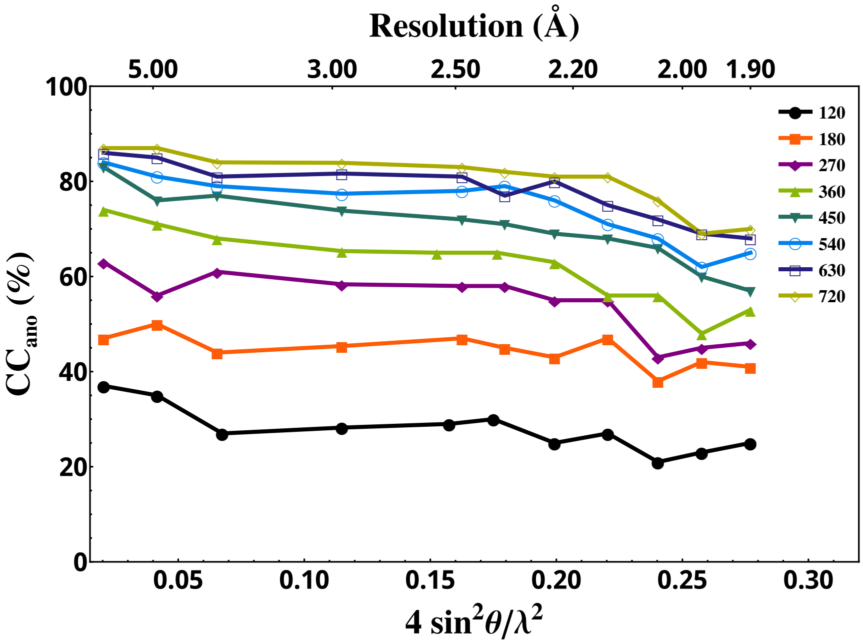


Figure III.1.8: CC$_{\text{ano}}$ as a function of resolution for the 8 wedges of data with different multiplicities

III.1.2.2 Effect on sub-structure determination

The first step in applying P-SAD to solve nucleic acids structure is to determine the phosphorus sub-structure using anomalous differences. To investigate the effect of data multiplicity on phosphorus sub-structure determination, 1000 trials of sub-structure search in *SHELXD*[52] were performed with SRL data at different multiplicities (Table III.1.4). The *SHELXD* solutions at different multiplicities were analyzed with CC$_{\text{all}}$ vs. CC$_{\text{weak}}$ figures (Figure III.1.9). Each dot in the figure represented a sub-structure search trial. The separation of the clusters in CC$_{\text{all}}$ vs. CC$_{\text{weak}}$ plot (section II.3.2.1) indicated the correct solutions (obtained for data sets **540**, **630** and **720**, see Figure III.1.9).

For data with lower multiplicity, only random solutions were obtained. With higher multiplicity, the success rates of sub-structure determination was increased. Clear solutions appeared when data multiplicity reached 10.9 (data **540**) and were further improved with increased multiplicity (Table III.1.5).

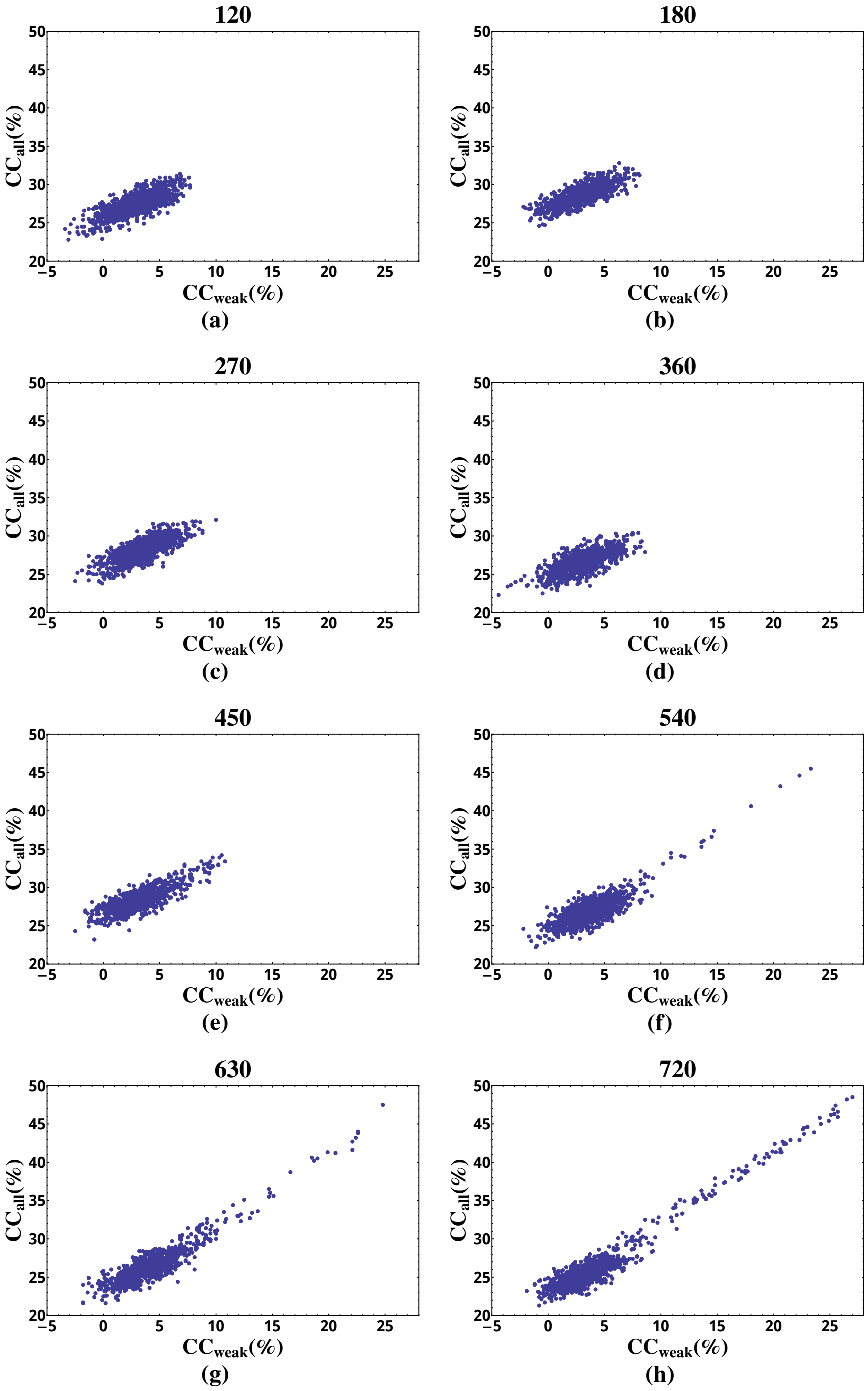


Figure III.1.9: CC_{all} vs. CC_{weak} profiles for data (a) **120**, (b) **180**, (c) **270**, (d) **360**, (e) **450**, (f) **540**, (g) **630**, (h) **720**.

Table III.1.5: The numbers of correct sites in *SHELXD* solutions for data with different multiplicities

Data	720	630	540	450	360	270	180	120
Multiplicity	14.5	12.7	10.9	9.1	7.3	5.3	3.6	2.5
Number of sites	26	22	18	3	1	2	3	2

III.1.2.3 Effect on P-SAD phasing

The next step after sub-structure determination is phasing. To evaluate the effect of data multiplicity on P-SAD phasing, phases and density maps obtained with data at different multiplicities were compared. The phases were calculated with known SRL phosphorus structure in *PHASER* as described in section II.3.2.2.

Phase quality was evaluated by calculating the phase errors between the experimental phases and the final refined calculated phases (Figure III.1.10). Phases calculated from high multiplicity data were shown to be more accurate than that of low multiplicity data. The phase errors for high multiplicity data (**360** to **720**), was below 40° corresponding to map CC higher than 75%. For low multiplicity data (**120**, **180**), phase errors after density modification were about 50° to 60° corresponding to map CC of 64% to 50%. Improvement in map quality was observed with increased data multiplicity (Figure III.1.12). This was confirmed by the map correlation coefficient (map CC) between the experimental phasing map and final refined map (Figure III.1.11).

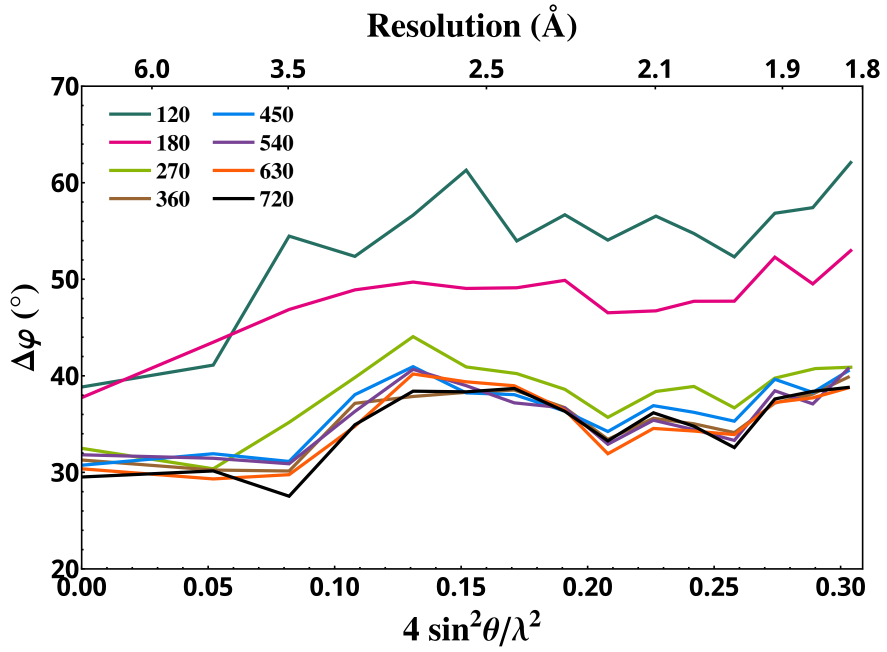


Figure III.1.10: Phase differences ($\Delta\varphi$ in degree) between the SAD phases against and phases calculated from refined model

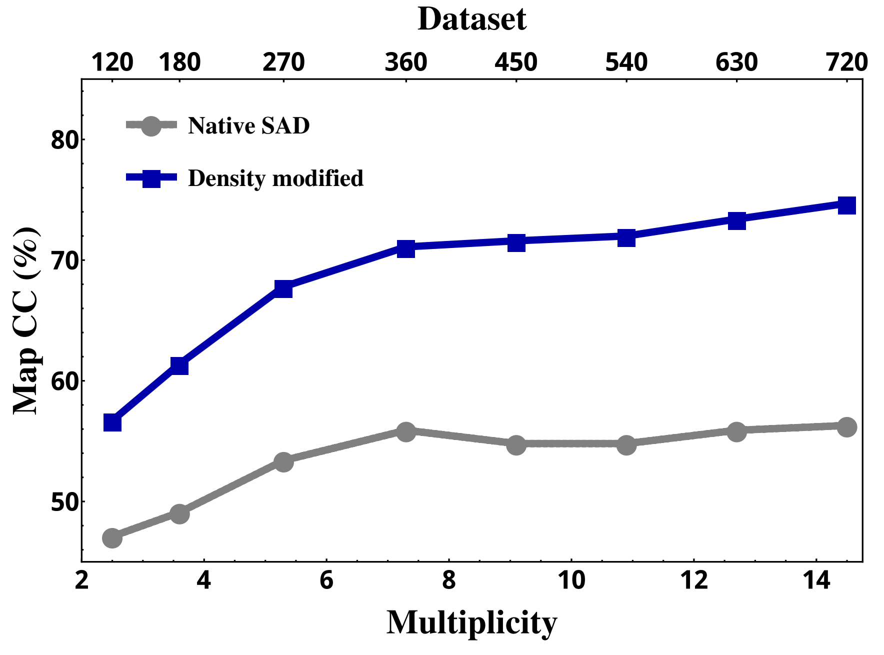
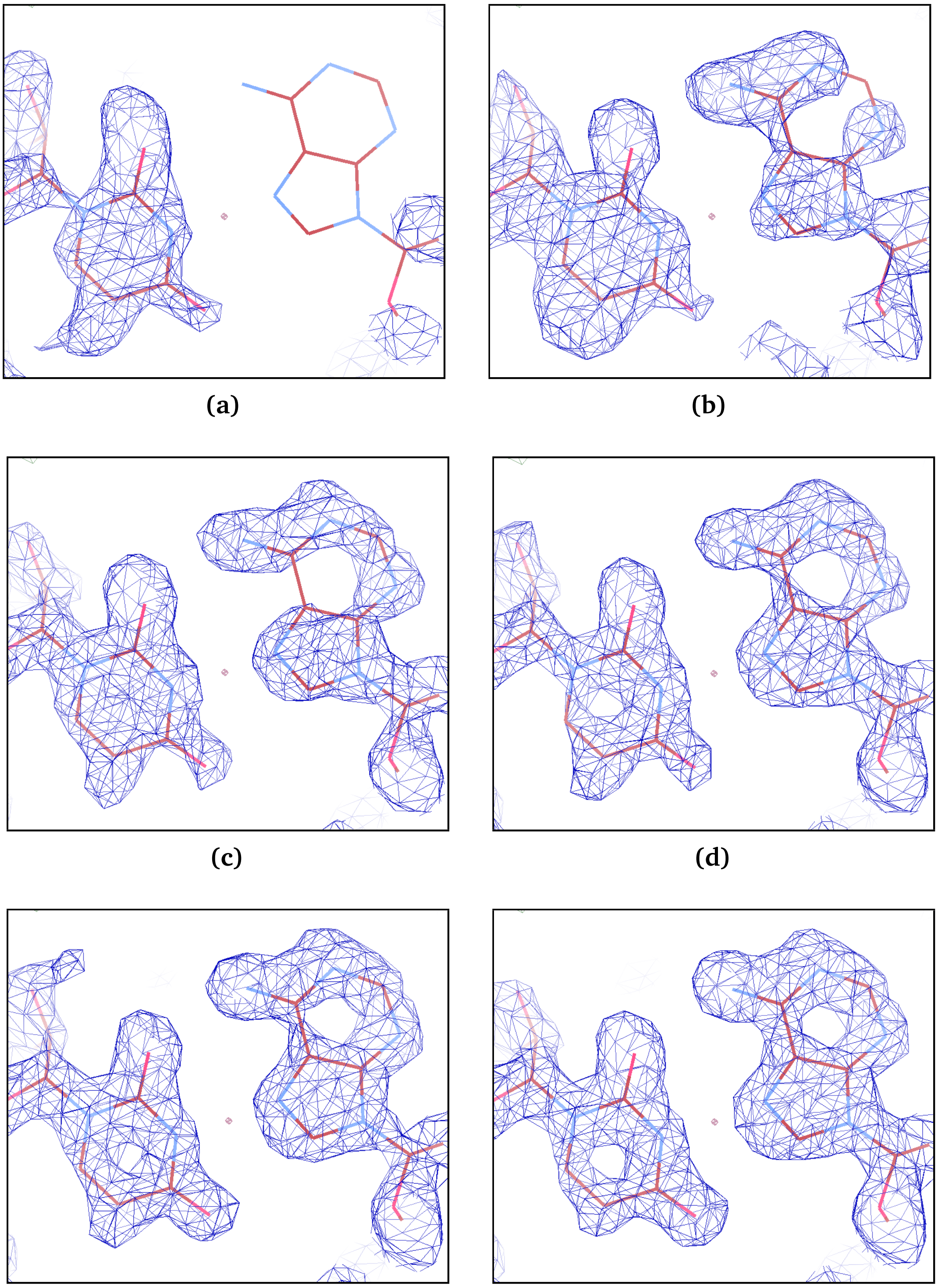


Figure III.1.11: Map CC of the experimental phasing before (gray) and after density modification (blue) for SRL data with different multiplicities



(a) (b) (c) (d) (e) (f)

Figure III.1.12: Experimental density map of data sets at different multiplicities: (a) **120**, (b) **180**, (c) **270**, (d) **360**, (e) **540**, and (f) **720**. The maps were obtained after density modification with *DM* and contoured at 1.5 σ.

III.1.3 Applying P-SAD to novel structures

Two novel structures including an octameric RNA duplex and a protein-DNA complex (σ^{E2}–TGTCAAA) were successfully determined by P/native-SAD. High multiplicity data were collected in both cases. Flux and exposure times were adjusted to control the radiation damage. Fine φ-slicing was used to improve the intensity to noise ratio for weak reflections. Multi-orientation data collection was also used to reduce systematic errors. Data statistics are shown in section II.3.1. The following sections will show the results of sub-structure determination and phasing of these two targets.

III.1.3.1 Octameric RNA duplex

The octameric RNA duplex containing 14 phosphorus (section II.1.1) diffracted to 1.7 Å (section II.3.1). The sub-structure was successfully determined in *SHELXD*.[52,104] The CC_{all} vs. CC_{weak} plot showed that the solutions were obtained after 1000 trials in *SHELXD* (Figure III.1.13). With the determined sub-structure, P-SAD phasing was successful at resolution 2.0 Å. An interpretable electron density map was obtained after density modification (Figure III.1.14). The map CC between the experimental phasing map and the map calculated from the refined model was 82.2%.

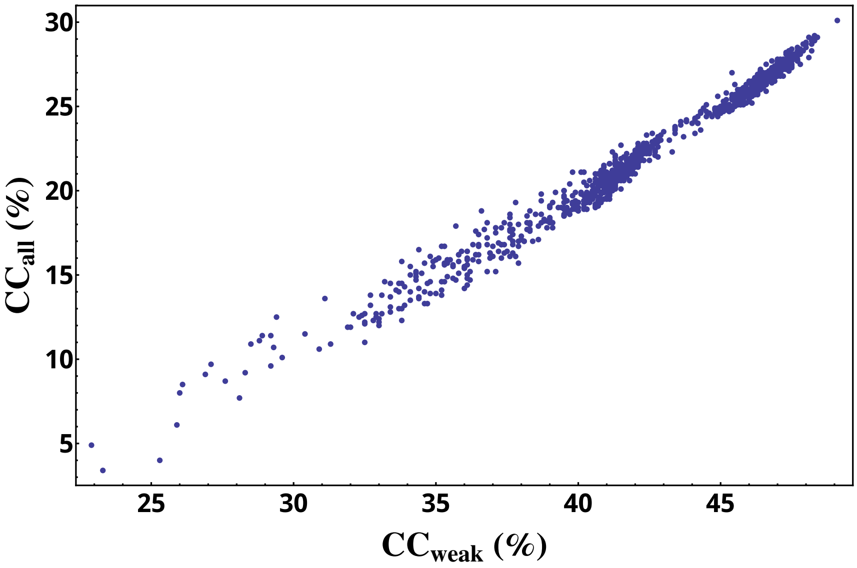


Figure III.1.13: CC_{all} vs. CC_{weak} plot of 1000 *SHELXD* trials sub-structure search for the octameric RNA duplex. Separations of the points in clusters indicated that the correct solutions were obtained.

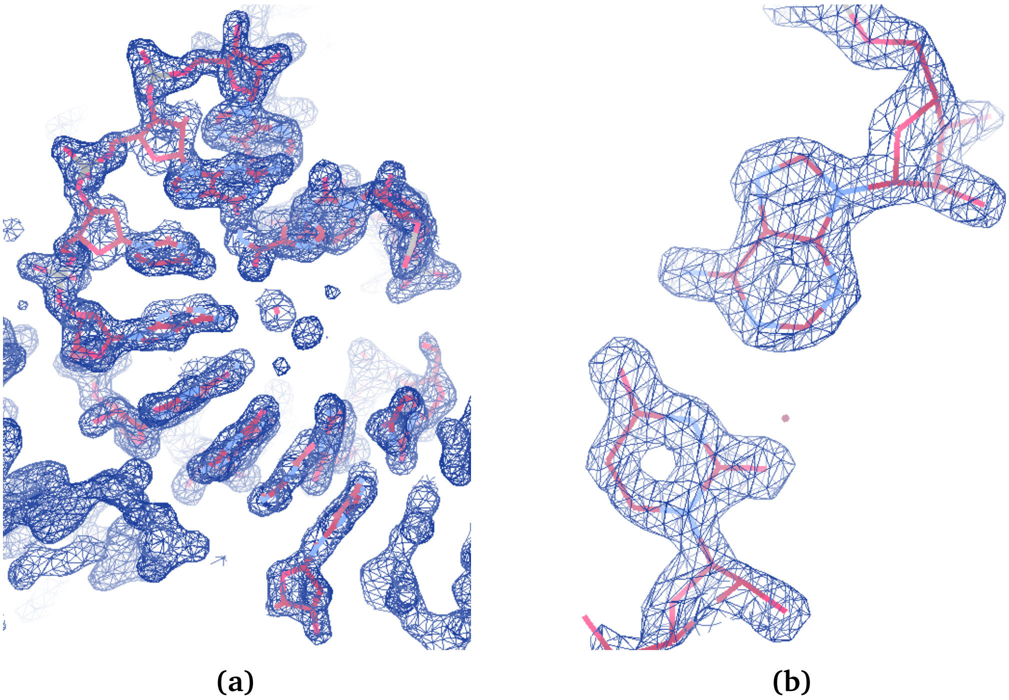


Figure III.1.14: Electron density map after density modification out of *PHASER* + *DM* for the octameric RNA duplex (obtained with weighted structure factors FDM and density modified phases PHIDM out of *DM*). (a) Section encompassing the molecule and fitting the refined model. (b) Close-up of an A-U nucleobase pair. The maps were contoured at 1.5 σ level.

III.1.3.2 σ^{E2}–TGTCAAA protein-DNA complex

The structure of a σ^{E2}–TGTCAAA protein-DNA complex (section II.1.1) containing 212 amino acids and 14 nucleotides was succesffuly determined with P-SAD.[94] The crystal was collected at a wavelength of 2.0 Å and diffracted to resolution 2.1 Å (Table II.3.1). At resolution cut-off of 2.5 Å, clear solutions were obtained with 1000 trials of *SHELXD* search (Figure III.1.15).

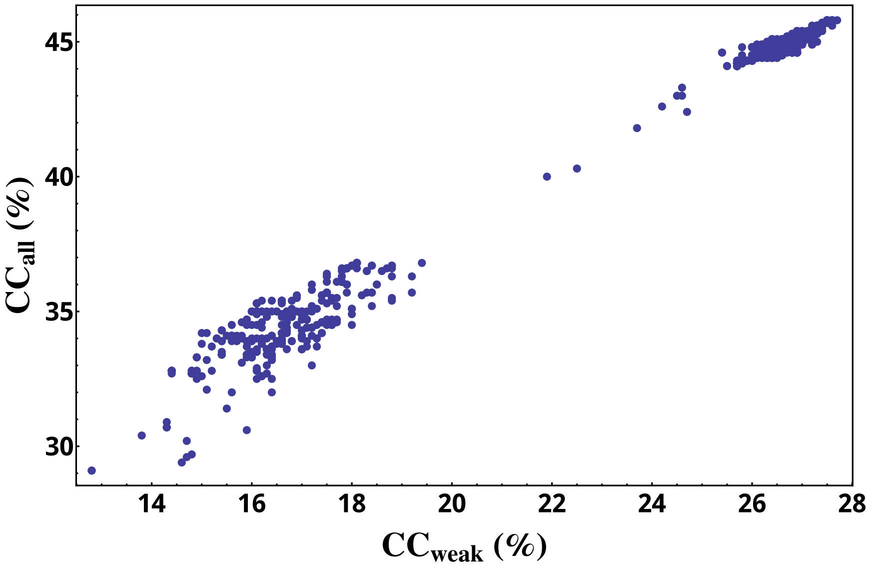


Figure III.1.15: CC_{all} vs. CC_{weak} plot of 1000 *SHELXD* trials sub-structure search for the protein-DNA target. Two separate clusters indicated that the correct solutions were obtained.

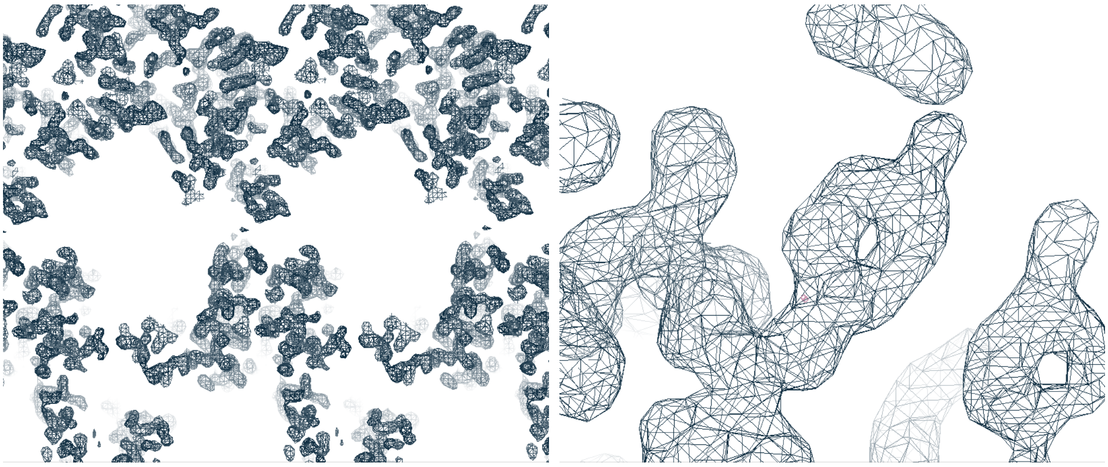
To find the resolution limit, sub-structure determination with *SHELXD* was performed at various resolution cut-off. Solutions were found until 4.6 Å (Table III.1.6). With the determined sub-structure, the whole complex structure could be determined with P-SAD phasing. Interpretable density map was obtained after density modification (Figure III.1.16). The map CC between the experimental phasing map and the map calculated from refined model was 81.3%.

---

[94] S. Campagne et al. *Nat. Struct. Mol. Biol.*, **21**: 269–276, 2014.

Table III.1.6: Number of correct sites out of sub-structure determination with *SHELXD* at resolution cut-off from 2.2 to 5.5 Å

Resolution (Å)	2.2	2.7	3.2	3.7	4.6	4.7	4.9	5.0	5.5
Number of sites	12	12	10	10	8	2	0	1	0



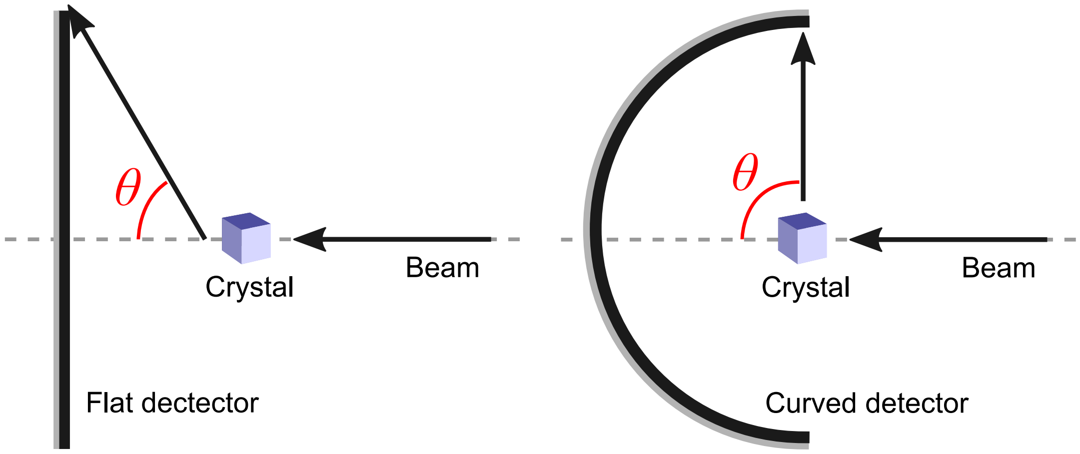
(a) **(b)**

Figure III.1.16: Electron density map after density modification out of *PHASER* + *DM* for the protein-DNA structure (obtained with weighted structure factors FDM and density modified phases PHIDM out of *DM*). (a) Section encompassing the molecules and showing a clear solvent boundary. (b) Close-up of a tyrosine side chain with a hole in the benzene. The maps were contoured at 1.5 σ level.

III.1.4 Summary and discussion

High multiplicity data collection was found to improve data quality and was beneficial to both phosphorus sub-structure determination and P-SAD phasing. However, accumulated X-ray dose introduces radiation damage. To control the X-ray dose, the flux and exposure were adjusted for high multiplicity data collection. Fine φ-slicing and low dose data collection can better sample the integrated intensities and control radiation damage at the same time. For the high multiplicity measurement, a multi-orientation goniometer, used to orientate the crystal differently, helps to reduce systematic errors.

Enhancement of the anomalous signal was observed for long wavelength data. However, due to the absorption effect and hardware limitation (limited detector distance and size), measurable resolution can be limited. As P-SAD sub-structure determination and phasing required high resolution data (section III.1.1.2, III.1.1.3), 1.6 Å wavelength was chosen to compromise the gain in anomalous strength with the loss of high resolution data. To really benefit from the improvement of anomalous strength with long wavelength, different detector geometry (2-theta or curved detector, see Figure III.1.17) would be beneficial to collect higher angle diffraction. Considering the air absorption, a helium filled path could also be used to further improve measurement accuracy in the future.



(a) **(b)**

Figure III.1.17: Limitation of measurable diffraction angles for (a) Flat detector and (b) Curved detector

Based on these discoveries, we implemented a data collection strategy for P-SAD combining high multiplicity, low dose, fine φ-slicing, and multi-orientation towards the goal of obtaining very accurate data. High multiplicity can help to reduce statistical errors and improve data quality. With increased multiplicity, and to limit the radiation damage, data have to be collected with proper flux and exposure time. *PILATUS* detector allows fine φ-slicing data collection which improves signal to noise ratio for weak reflections. With the multi-orientation goniometer PRIGo, crystals can be orientated differently during data collections. This helps to avoid badly diffracted domains of the crystal and reduce systematic errors in high multiplicity data collection. For certain space group, the crystal can be orientated in a way that *Bijvoet* pairs are collected on a same image, *i.e.*, with the same X-ray dose. This is an elegant way of reducing systematic errors in anomalous differences.

The average number of observations per phosphorus site in the asymmetric unit (#Observation/#P) is an indicator for the success of P-SAD phasing. With the help of this indicator, we investigated the potential of applying P-SAD in solving protein-nucleic acids structures. The #Observation/#P ratio was calculated for the 4 different RNA targets (Table III.1.7). Values of #Observation/#P ratio in the range of 100 - 200 were found to be required to solve the phosphorus sub-structures. This means that in order to determine large RNA structures, large amount of observations are required.

Table III.1.7: The number of observations divided by the number of phosphorus sites (#Observation/#P) at resolution cut-off from 1.8 to 2.5 Å for 4 RNA targets. For the successful determined sub-structures the ratios are colored in green and the ratios for failed sub-structure determinations are in red.

(a) Experimental data

Molecule	Number of P	Resolution (Å)						
		1.8	2.0	2.1	2.2	2.3	2.4	2.5
2×8-mer	14	285	209	180	156	136	121	106
SRL	26	225	164	141	123	108	95	84
DIS	44	–	–	153	133	116	102	90
GIR1	186	–	–	–	–	–	101	90

(b) Simulated data

Molecule	Number of P	Resolution (Å)						
		1.8	2.0	2.1	2.2	2.3	2.7	2.8
2×8-mer	14	287	211	182	159	138	86	77
SRL	26	228	166	143	124	109	67	61
DIS	44	247	179	153	133	116	70	63
GIR1	186	246	178	153	133	116	70	63

To extend sub-structure solution to lower resolution range, larger #Observation/#P ratio was required. This can be achieved by reducing the number of phosphorus within a molecule. Simulated data of DIS RNA with reduced number of phosphorus were used to test this idea (Table III.1.8). We observed that by artificially reducing the number of phosphorus in the structure, and therefore increasing the #Observation/#P ratio, successful sub-structure solution could be obtained at much lower resolution.

Table III.1.8: #Observation/#P ratio at different resolution cut-off for DIS RNA data simulated with reduced number of phosphorus atoms. Colored numbers are the #Observation/#P ratio at the corresponding resolution cut-off. The ratios for the successful determined sub-structures are colored in green. The ratios for the failed sub-structure determinations are in red.

Number of P	Resolution (Å)									
	2.0	2.1	2.3	2.6	2.8	3.0	3.1	3.5	4.0	4.1
5	1532	1314	991	676	533	429	387	262	172	159
10	766	657	495	338	266	215	194	131	86	80
15	510	438	330	225	177	143	129	87	57	53
20	383	328	247	169	133	108	97	66	43	40
25	306	262	198	135	106	86	77	52	34	32
30	255	219	165	112	89	72	65	43	29	27
35	218	187	141	96	76	61	55	38	25	23
40	191	164	124	84	66	54	48	33	22	20
44	179	153	116	80	63	51	46	31	21	19

It is worth to note that the anomalous peak heights measured in σ map are increased with lower phosphorus content. However, this did not reflect a real improvement in the overall anomalous signal to noise ratio. Statistically, standard deviation of measurements was inversely proportional to the square root of the amount of measurements. The standard deviation of density distribution over all grids was given by,

$$\sigma_\rho = \sqrt{\frac{1}{N_g}\sum_{i}^{N_g}(\rho_i - \overline{\rho})^2} \qquad \text{(III.1.4.1)}$$

where N_g denotes for the total number of grids composed the density map and ρ_i refers to the density at each grid and $\overline{\rho}$ is the mean value of density distribution over all grids in the anomalous map. Assuming that the noise level was uniform over the whole map grids, that the anomalous peaks were significantly stood out from the background and that the contribution of each anomalous scatters was even, then,

$$\begin{aligned}\overline{\rho} &= \frac{1}{N_g}\sum_{i}^{N_g}\rho_i \\ &\approxeq \frac{1}{N_g}\left(\sum^{N_p}\rho_p + \sum^{N_n}\rho_n\right) \\ &\approxeq \frac{1}{N_g}\left(N_p\rho_p + N_n\rho_n\right) \end{aligned} \tag{III.1.4.2}$$

where ρ_p and ρ_n are the density measured at grids for peak and noise respectively and N_p and N_n are the number of grids counted for the peaks and background noise respectively. The relation between N_g, N_p and N_n is $N_g = N_p + N_n$. Plugging equation III.1.4.2 in equation III.1.4.1 gives,

$$\sigma_\rho \approxeq \sqrt{\frac{(N_g - N_p)N_p}{N_g^2}\left(\rho_p - \rho_n\right)^2} \tag{III.1.4.3}$$

Normally there are only few anomalous scatters in the structure and thus the number of grids constructing the anomalous peaks in the density maps is much less than the grids number of the whole map. With assumption of $N_p \ll N_g$, equation III.1.4.3 can be simplified as,

$$\sigma_\rho \approxeq \sqrt{\frac{N_p}{N_g}\left(\rho_p - \rho_n\right)^2} \propto \sqrt{N_p} \tag{III.1.4.4}$$

Therefore the σ used to measure the peak heights in the anomalous density map were related to the amount of anomalous scatters with a square root function. With decreasing number of phosphorus atoms, the σ level of the overall map was decreased by square root of N_p. Therefore, the value of peak heights in the unit of σ increased by square root of N_p. This relationship between the anomalous peak heights and the number of anomalous scatters was clearly illustrated by the fit between the calculated curve of $\sqrt{N_p}$ and the experimental measured peak heights in the unit of σ (Figure III.1.18). In summary, the increase of the peak heights is due to the decrease of σ with lower amount of sites.

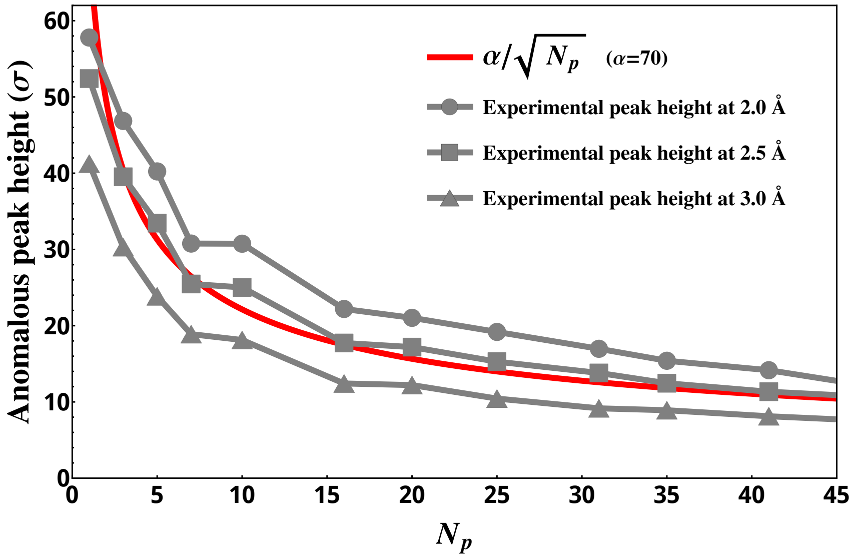


Figure III.1.18: The anomalous peak height in unit of σ (marked grey) and the calculated curve of square root of the reduced number of phosphorus ($\alpha/\sqrt{N_p}$ in red) in the DIS structure. The anomalous peak heights were calculated with the Fourier synthesis of the simulated anomalous differences and phases from the reduced DIS RNA models.

Protein-DNA or Protein-RNA complexes do naturally have a lower phosphorus content, and therefore have a higher #Observation/#P ratio compared to DNA or RNA only (Table III.1.9). This explains the success of sub-structure determination for the protein-DNA σ^{E2} – TGTCAAA complex at lower resolution. While the resolution required for the success of sub-structure determination can be quite different for each target, the #Observation/#P ratio requirements can be used as a general parameter to predict the success of sub-structure determination for different types of macromolecules containing nucleic acids.

Table III.1.9: #Observation/#P ratio at resolution cut-off from 1.8 to 2.5 Å for 4 RNA targets and the protein-DNA complex. For the successful determined sub-structures the ratios are colored in green and the ratios for the failed sub-structure determinations are in red.

Molecule	Number of P	Resolution (Å)					
		2.2	2.3	2.4	2.5	4.6	4.7
Complex	12	652.7	618.5	573.9	520.8	85.2	79.5
2×8-mer	14	156	136	121	106	16.5	15.4
SRL	26	123	108	95	84	13.1	12.2
DIS	44	133	116	102	90	13.1	12.4
GIR1	186	–	–	101	90	13.6	12.6

Chapter III.2

Pushing the limit of P-SAD: towards *de novo* MR-P-SAD

As presented in section III.1.1 and III.1.2, highly accurate and high resolution data are required in order to apply P-SAD. For larger RNA structure, due to the intrinsic limitation in its sub-structure complexity, only very low #Observation/#P ratio can be obtained (section III.1.4). This restricts the applicability of P-SAD to solve large nucleic acid structures.

When an analogous molecule or domain is already known, Molecular Replacement (MR)[126] an be used in structure determination to complement P-SAD. But the efficiency of MR highly depends on the quality of the search model. As MR can provide additional phase information, it can be of interest to combine MR with P-SAD when dealing with large nucleic acids and complexes structures.

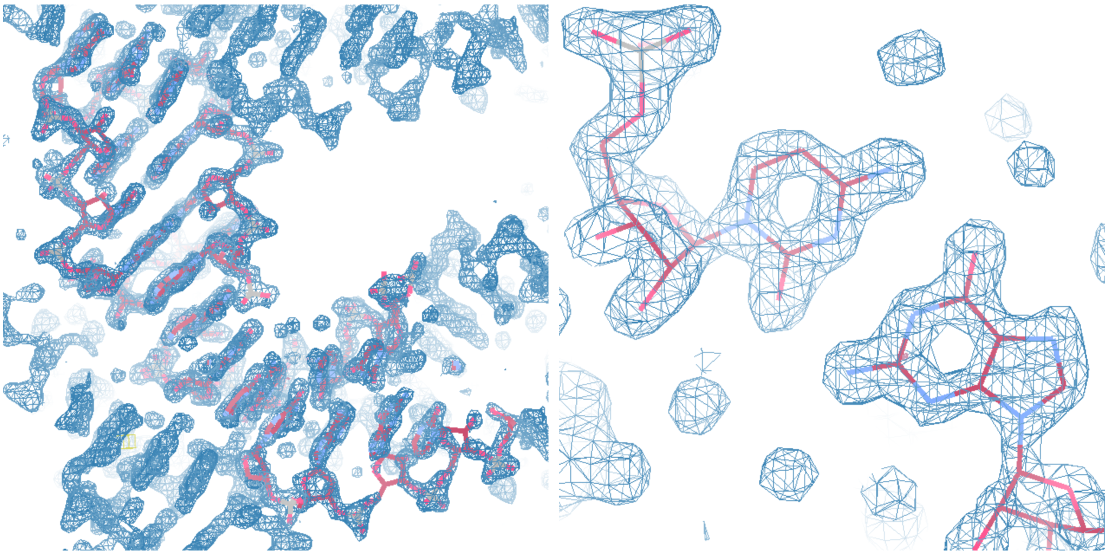
---

[126] M. G. Rossmann. *Acta Crystallogr. Sect. A Found. Crystallogr.*, **46**: 73–82, 1990.

III.2.1 MR with full model

III.2.1.1 MR with true model

In order to test the performance of Molecular Replacement for the RNA targets, the SRL model was used as the search model in *PHASER*.[89] The experimental data were collected at wavelength of 1.6 Å and diffracted to 1.6 Å. The high translation function Z-score of 33.1 and log-likelihood gain of 1320 out of *PHASER* indicated the correct solution. The electron density map $(2m|F_{obs}| - D|F_{calc}|\exp(i\alpha_{calc}))$ out of *PHASER* was readily interpretable. This test showed that, with the true full search model, Molecular Replacement was feasible to determine the SRL RNA structure.



(a) (b)

Figure III.2.1: The electron density map $(2m|F_{obs}| - D|F_{calc}|\exp(i\alpha_{calc}))$ of the SRL structure obtained from MR in *PHASER*. (a) Section encompassing the molecule and fitting the refined model. (b) Close-up of a G-C nucleobase pair. The maps were contoured at 1.5 σ level.

---

[89] A. J. McCoy et al. *J. Appl. Crystallogr.*, **40**: 658–674, 2007.

III.2.1.2 MR with distorted model

In order to test the dependency of the success of MR search on the accuracy of search model, shifts of the atomic coordinates were added in SRL RNA model using normal mode perturbation[127,128] with a self-written *Mathematica* script.[117,118] The root mean square deviation (RMSD) between the distorted model to the true refined model was increased from 0.5 to 3.0 Å by step of 0.5 Å. (Figure III.2.2).

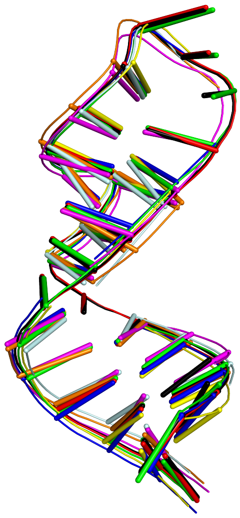


Figure III.2.2: Distorted SRL models obtained with normal mode perturbation. Error were added with root-mean square deviation (RMSD) to true target structure from 0.5 to 3.0 Å.

Together with the original SRL structure, a total of 7 models were used as the initial models in Molecular Replacement using *PHASER*. The data were truncated to 2.5 Å for each MR search job. The translation functions Z-score (TFZ, section II.3.3.1) of the MR search at different RMSD levels are showed in Figure III.2.3. At RMSD smaller than 2.0 Å, high TFZ scores (> 8) indicate the correct solutions. When RMSD is above 2.0 Å, no solutions were obtained from *PHASER*. This illustrated that high structural similarity (RMSD < 2.0Å) was required for the success of MR search.

---

[127] M. Delarue and P. Dumas. *Proc. Natl. Acad. Sci. U. S. A.*, **101**: 6957–62, 2004.
[128] K. Suhre and Y. H. Sanejouand. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **60**: 796–799, 2004.
[117] Wolfram Research Inc. *Mathematica Edition: Version 9.0* Champaign, Illinois, 2012
[118] N. Ambert et al. *CCP4 Newsl.*, **44**: 22–31, 2006.

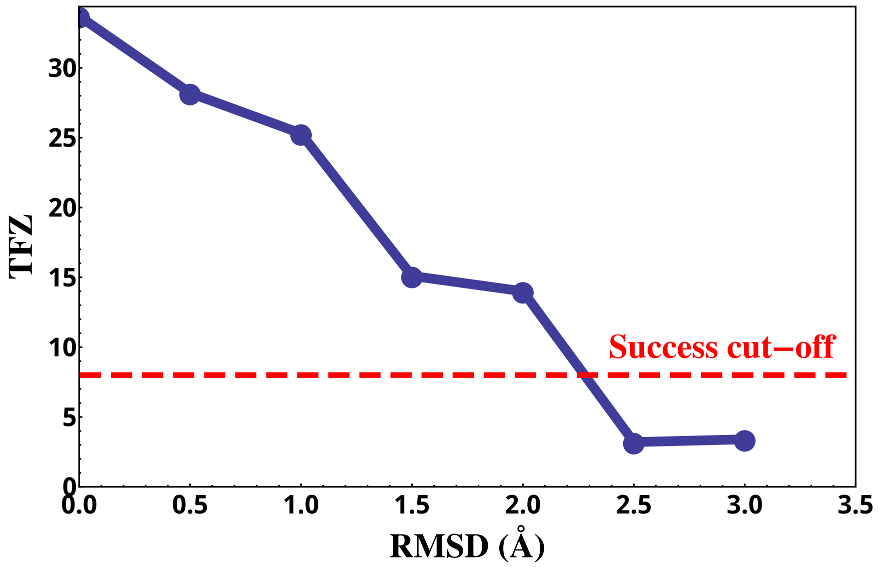


Figure III.2.3: The translation function Z-scores (TFZ) out of *PHASER* plotted against the RMSD of search models. The data used in MR search were truncated to 2.5

III.2.2 MR with model fragments

Full homologous structure are not usually available. But partial structures can be used for MR.[129] In some cases, the partial solution can be completed via several round of model completion plus additional MR search and refinement.[130] RNA structures have abundant structural motifs (section I.1.2.2), which can be used as a starting point for solving the whole structure with MR. This section will study how the partial model and model fragment library can be used in MR. The first part deals with partial model fragments extracted from true model, then in the second part we evaluate the requirement of model accuracy for MR using model fragments and finally the ideal fragments library is tested in MR search.

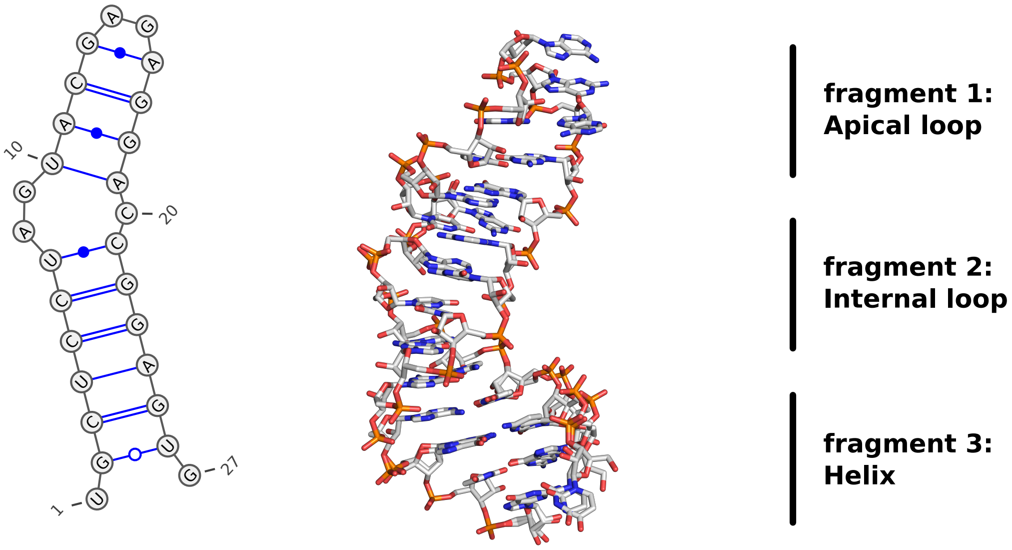
III.2.2.1 MR with true model fragments

- **SRL RNA model**

---

[129] A. Laurenzi et al. *Int. J. Mol. Sci.*, **14**: 14892–907, 2013.
[130] G. Scapin. *Acta Crystallogr. Sect. D, Biol. Crystallogr.*, **69**: 2266–75, 2013.

To test the performance of MR with RNA model fragments, fragments of the SRL structure (PDB code: 1Q9A,[121] 27 nucleotides) were extracted from the true model. Each fragment had 25 to 30% sequence identity to the full sequence (Figure III.2.4). Each model fragment was displaced with random rotation and translation to simulate the randomized starting point for a MR search. The diffraction data used in the MR search were truncated to resolution 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.5 and 6.5 Å.



(a) **(b)**

Figure III.2.4: (a) The 2D illustration of the secondary structure of SRL RNA, (b) The 3D structure of the SRL RNA

The maximum likelihood phasing methods in *PHASER* requires prior knowledge of the deviation of the search models from the real structure.[89] Therefore it was important to have a look at the model B-factors before running MR. In the SRL structure the helix region had higher order of flexibility and thus relatively higher B-factors was observed (Figure III.2.5). The higher B-factors indicated higher level of model errors which may influence the MR search. To achieve better results, adjustment in resolution cut-off was required.[116]

The results of MR search were evaluated by the translation function Z-score

---

[121] C. C. Correll et al. *Nucleic Acids Res.*, **31**: 6806–6818, 2003.

[116] A. J. McCoy. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **63**: 32–41, 2007.

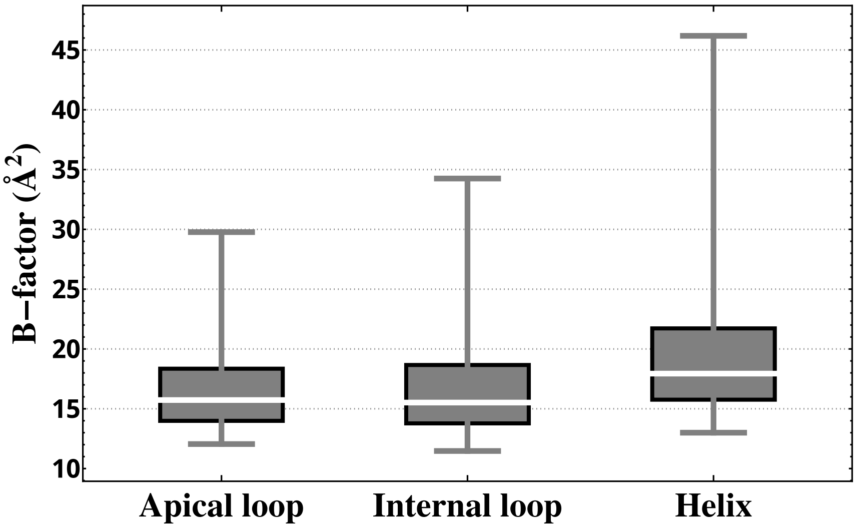


Figure III.2.5: B-factors of the 3 model fragments of SRL RNA

(TFZ score, section II.3.3.1). Higher than 8 TFZ score indicates the correct solution. As shown in Figure III.2.6, solutions were found at resolution above 5.5 Å for the apical loop and internal loop fragments and correct solutions appeared at resolution above 4.5 Å for the helix fragment. The latter one (helix) had lower RMSD compared to that of apical loop and internal loop. This was due to the effect of larger model errors of the helix fragment represented by higher B-factors shown in Figure III.2.5.

MR solutions and density maps were checked in *COOT*.[113] The maps were found to be interpretable. Due to the missing part of the structure, the complete RNA molecule were not visible in the density map (Figure III.2.7a). However, the region closed to the search fragments showed interpreatable features (Figure III.2.7b). With the help of known sequence of the RNA chain, the missing part of the structures could be modeled in the density. Then it was possible to run a new cycle of MR search/building to complete the structure.[119]

---

[113] P. Emsley and K. Cowtan. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **60**: 2126–2132, 2004.

[119] M. P. Robertson et al. *Methods*, **52**: 168–172, 2010.

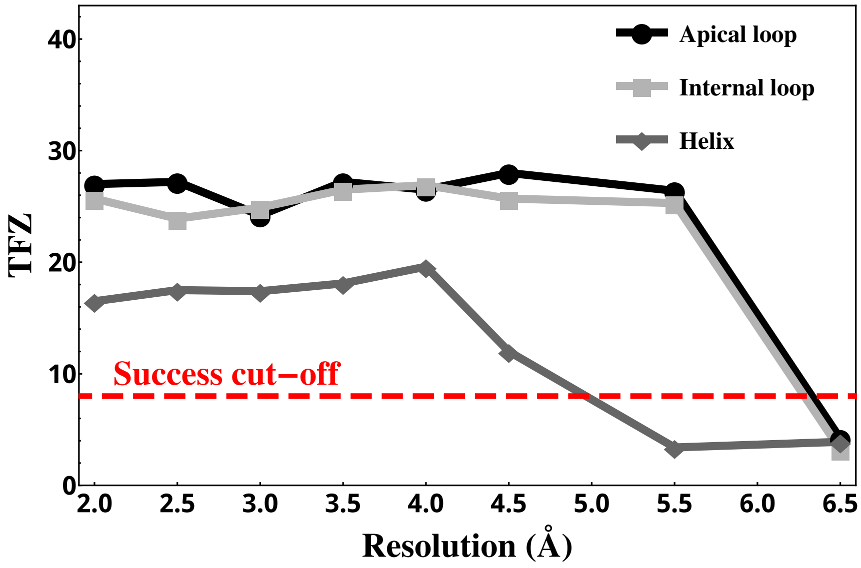
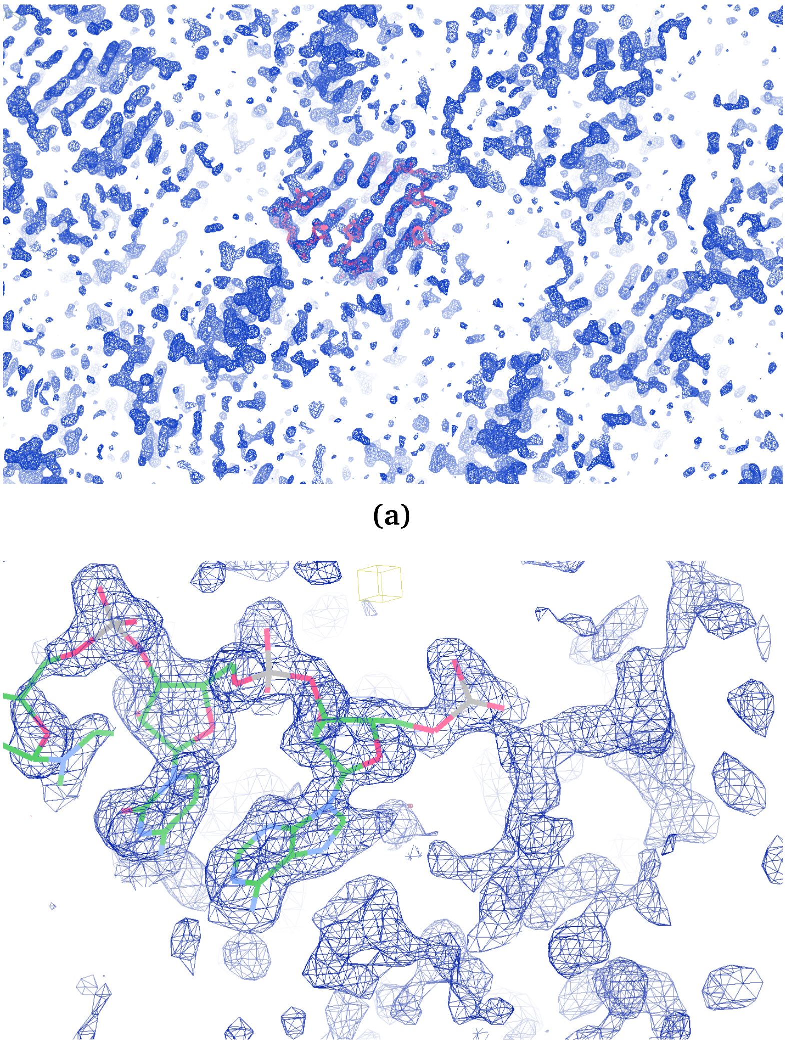


Figure III.2.6: Translation function Z-score (TFZ) of MR search with 3 partial SRL model fragments at different resolution cut-off.



(a)

(b)

Figure III.2.7: The electron density map $(2m|F_{obs}| - D|F_{calc}|\exp(i\alpha_{calc}))$ out of MR in *PHASER* with a model fragment (internal loop) extracted from SRL structure. (a) Overall map quality. (b) Electron density at the missing region of molecule showed interpretable coverage of the sugar-phosphates backbones and an Adenine nucleobase. The maps were contoured at 1.5 σ level.

According to M. P. Robertson *et al.*,[119] phases obtained from MR using partial model fragments can be extended through iterations of manually editing and MR search and refinement until completion of the model. A simplified strategy is shown in Figure III.2.8. In the case of MR with SRL model fragments, additional model fragments could be built into the density close to terminal residues as showed in Figure III.2.7b. The solution then can be completed through iterated MR with succeeding model editing and refinement as described above.

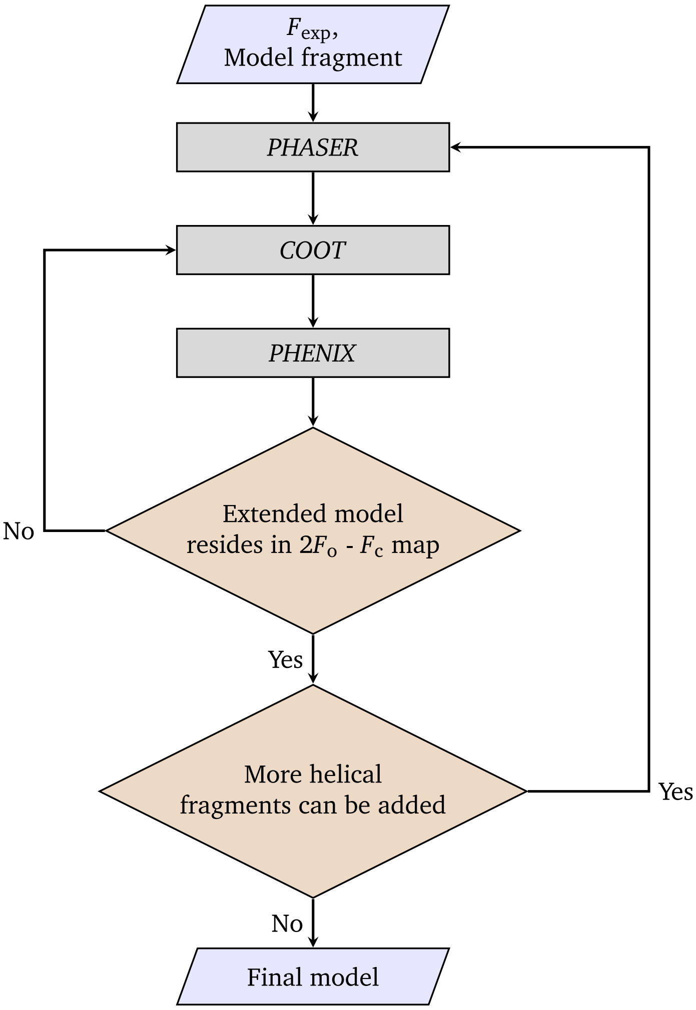


Figure III.2.8: Schematic of iterated MR search with model fragments. Initial MR search is performed in *PHASER* using experimental data and model fragments. The solution of MR search is examined and manually edited in *COOT* based on the σ_{A} -weighted $2F_o - F_c$ map. The extended model is then refined using *phenix.refine* in *PHENIX* package. If the σ_{A} -weighted $2F_o - F_c$ map out of refinement shows interpretable features in missing part. The model will be edited in *COOT* to add more helical fragments. Otherwise, unoccupied or wrongly modeled residues will be corrected and refined in iteration until the quality of the map is improved and interpretable density of missing region is found. The whole procedure can be repeated in iteration until no more fragments can be added and the output model is generated through a final refinement.

- **GIR1 RNA model**

To test the feasibility of using the MR fragments for solving large RNA structures, GIR1 RNA structure (188 nucleotides) was used as the target. 5 pieces of fragments were extracted from the refined structure and displaced in different orientation in the lattice (Figure III.2.9)

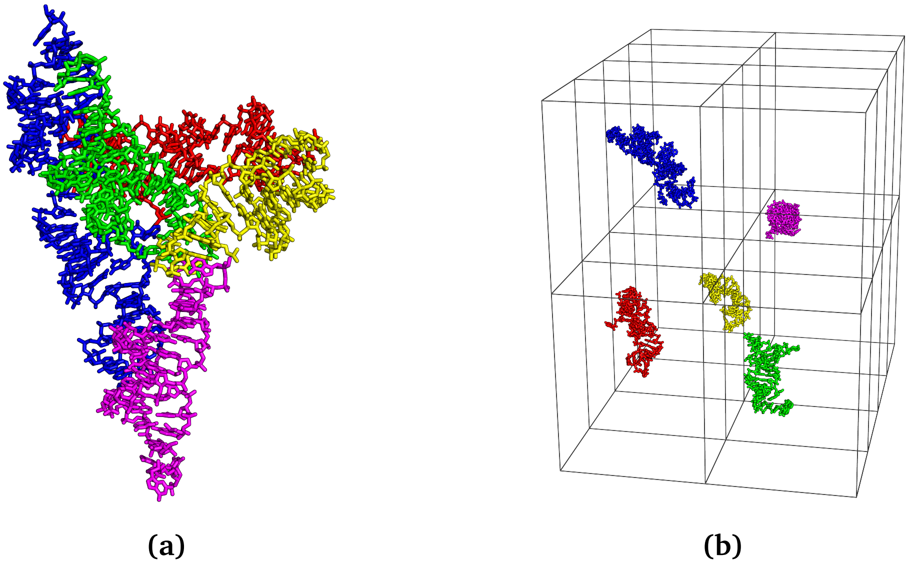


Figure III.2.9: The 5 pieces of fragments cut from refined GIR1 structure (a) 5 fragments showed in the GIR1 structure with different colors (b) Fragments displaced in different orientations to simulate the random starting point for MR search

The 5 fragments were then used as search model in *PHASER*. The solution from *PHASER* showed high TFZ score of 40.8 and log-likelihood gain of 1506.4. Four out of five fragments were correctly determined (Figure III.2.10). This experiment showed that it was possible to solve the GIR1 structure with MR using only small fragments. The iterate searching procedure was enabled in *PHASER*, in order to benefit from the determined fragments in each cycle of MR.

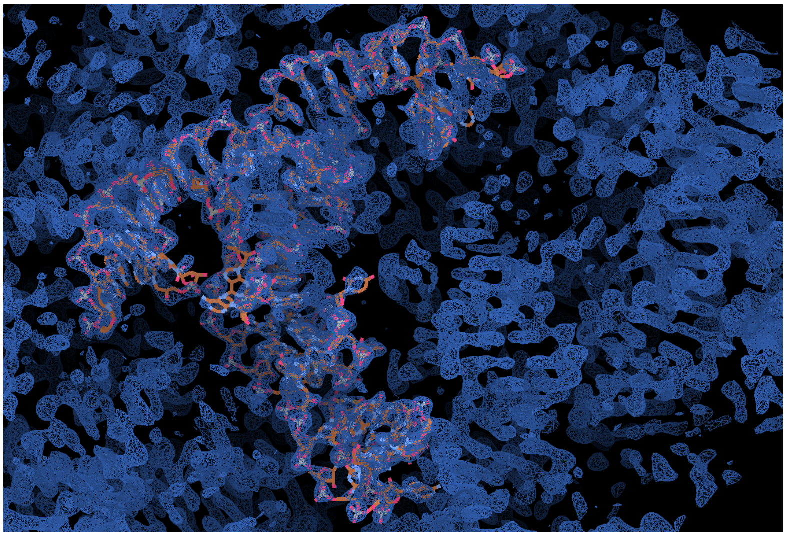
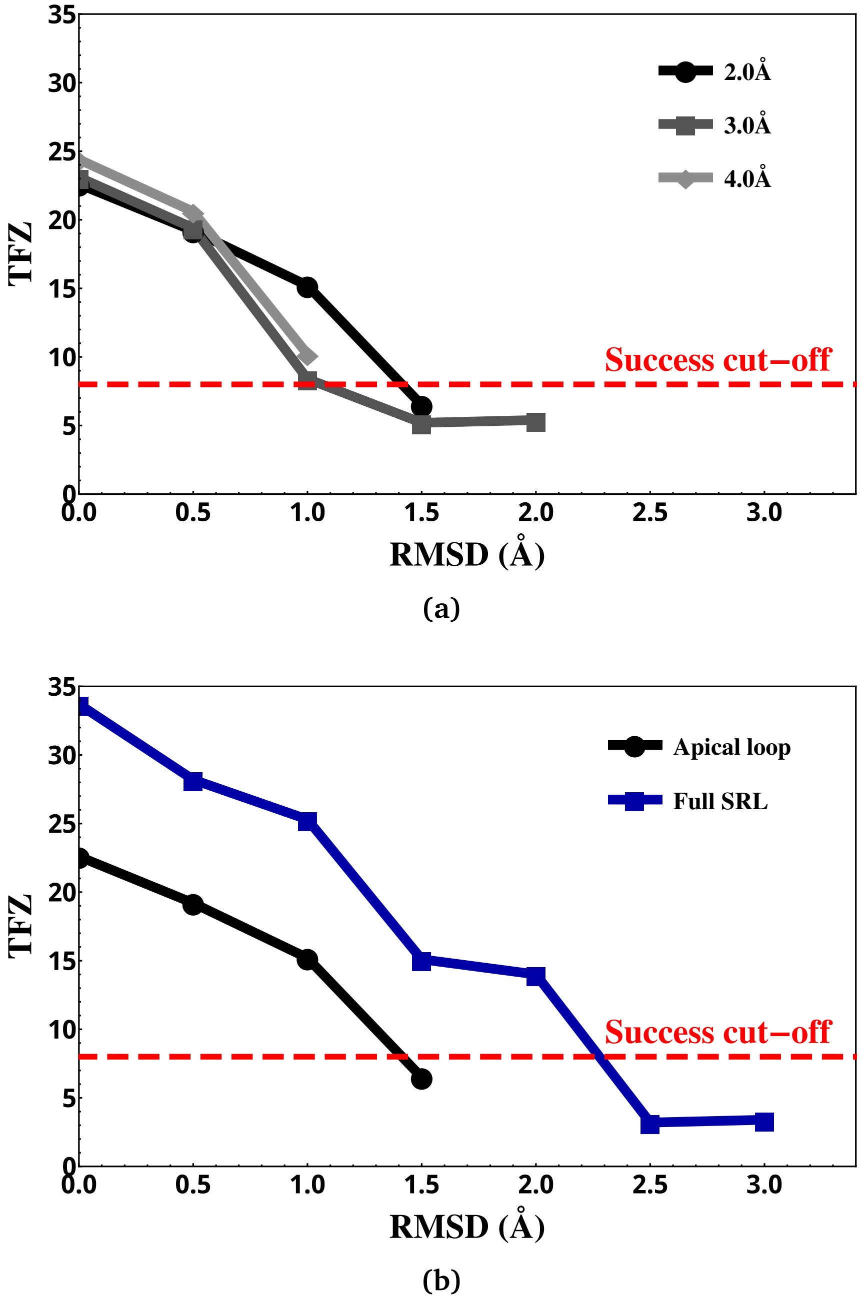


Figure III.2.10: The electron density map $(2m|F_{obs}| - D|F_{calc}|\exp(i\alpha_{calc}))$ out of MR in *PHASER* with model fragments of GIR1 RNA. The map was contoured at 1.5 σ level.

III.2.2.2 MR with distorted model fragments

The model fragments used in the above tests (section III.2.2.1) were extracted from the true SRL structure. In real life, the fragments obtained as initial search models are different from the true structure. In order to test the effect of coordinates errors in MR search with RNA fragments, an apical-loop fragment was extracted from true SRL model and coordinate shifts were added to its structure with normal model perturbation. The root-mean-square deviation (RMSD) of the distorted fragments were 0.5, 1.0, 1.5, 2.0 and 2.5 Å. The distorted apical-loops were used as the initial search models in MR search with *PHASER*. The data were truncated to resolution 2.0, 3.0 and 4.0 Å in MR search.

As shown in Figure III.2.11a, correct solutions were obtained when model RMSD was lower than 1.0 Å. With low resolution cut-off, the solutions did not get improved. Also, the overall TFZ score of MR solutions for the fragments was lower than that for the full search model (Figure III.2.11b). In the case of full model MR search, models with maximum 1.5 to 2.0 Å RMSD were allowed for successful MR search. For partial model MR, only 1.0 to 1.5 Å RMSD in the MR search fragments were tolerable.



(a)

(b)

Figure III.2.11: Translation function Z-score (TFZ) against the root-mean-square deviation (RMSD) of the distorted model. (a) TFZ out of MR at different resolution cut-off plotted against RMSD of the distorted apical-loop fragment. (b) Comparison of TFZ score of MR with the distorted full SRL model and the distorted apical-loop fragment.

III.2.2.3 MR with ideal model fragments

In sections III.2.2.1 and III.2.2.2, the model fragments used in MR search for SRL RNA comprise about 30% of the full sequence. The initial solution of the MR search with fragments was sufficient to initiate phase extension procedure via MR/model editing iteration and refinement. This approach could support structure determination when only part of the homologous structures is available. However, the model fragment used in the initial MR should have high structural similarity to the true structure. It was found that less than 1.5 Å RMSD of the search fragment was required for successful MR.

In practice, it is common that no homologous structure is available. In such cases, ideal model fragments can then be generated from known structures.[2] To test the concept of MR with ideal model fragments, 158 tetra-loop fragments were extracted from PDB[13] based on the secondary structure of SRL RNA with *ASSEMBLE*.[120] These fragments were then packaged in a home-made script for running MR search. The RMSD values between the top solutions (highest TFZ scores) and the true structure were calculated. PDB codes of the fragments and the corresponding TFZ scores and RMSD values of their MR outputs are shown in Table III.2.1.

---

[2] M. P. Robertson and W. G. Scott. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **D64**: 738–744, 2008.

[13] H. M. Berman et al. *Nucleic Acids Res.*, **28**: 235–242, 2000.

[120] F. Jossinet et al. *Bioinformatics*, **26**: 2057–2059, 2010.

Table III.2.1: Translation function Z-score (TFZ) and the root-mean-square deviation (RMSD) of the solution against true structure obtained from MR search with ideal SRL apical-loop fragments

Search model's PDB code	TFZ	RMSD (Å)
2HHH	5.7	1.5734
2QP0	7.2	1.9671
2I2U	3.4	3.0742
1I95	5.0	3.4536
3OR9	6.6	3.4930
1YL4	7.3	3.5060
1ZIG	3.2	3.8809
3I9D	6.1	4.0341

The solution with the lowest RMSD (2HHH) is shown together with the correct full model in the unit cell (Figure III.2.12). The rotation search was found to be correct as the solution was almost in the same orientation regarding the true structure. An obvious coordinates shift between the solution and the true model was observed. It increased the error between the solution and the true structure and generated a RMSD larger than 1.0 Å. Although the tetra-loop fragment generated a solution in almost the correct orientation, the TFZ score of this solution was not significantly high compared to the remaining solutions and no order or rank in TFZ scores was found among the solutions. Thus, the TFZ score alone is not sufficient to provide a proper indication of the correct solution. In the absence of the known structure, identification of the solution from noise is in fact difficult. In addition to the structural variation, the difficulty is the small size of the search fragment. Only 6 nucleotides were contained in the search model for this test, so that very limited phase information could be employed for the rotation and translation searches.

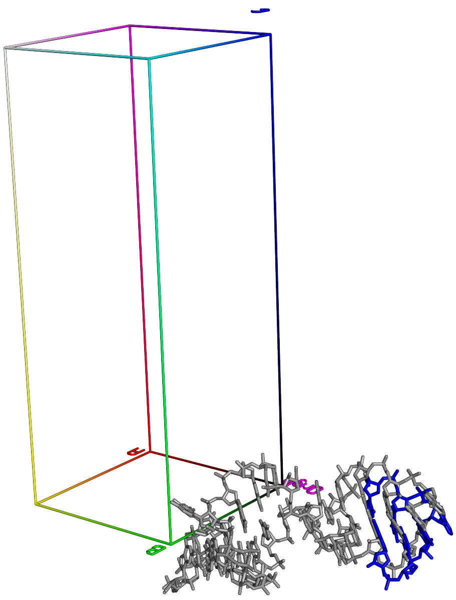


Figure III.2.12: The top solution of MR with tetra-loop fragment from structure 2HHH is showed together with the true SRL full model

III.2.3 Using anomalous signals in combination with MR

MR and experimental phasing with anomalous signals can be combined to tackle challenging situations (section I.3.3). In the following sections, two strategies of using anomalous signals in combination with MR are analyzed to determine nucleic acid structures. The first strategy consists in performing MR directly in anomalous differences with the phosphorus sub-structure. The second approach uses MR-P-SAD with partial models.

III.2.3.1 MR in anomalous differences

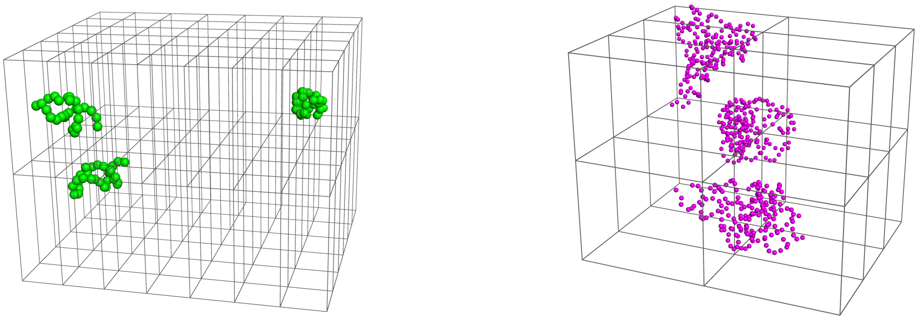
- **MR in anomalous differences with true P-model**

  The phosphorus sub-structure alone can be used in MR search against experimental anomalous differences instead of structure factors. In order to prove the concept and test the validity of this method, search models containing only phosphorus atoms were extracted from the true SRL RNA and GIR1 RNA structures. Three copies of random orientated and translated models were generated in *SUPERPOSE*[131] with SRL and GIR1 phosphorus model respectively. Then the three copies of phosphorus models were used as search model in *PHASER*[89] and the anomalous differences $|\Delta_{ano}|$ with the corresponding sigmas $|\sigma_{ano}|$ treated as normal structure factor amplitudes its standard deviations.

  The root-mean-square deviation (RMSD) of the solution was calculated against the true solution. Structures with high RMSD values ($>$ 1.0 Å) correspond to wrong rotation or translation searches in MR. Low RMSD values ($<$ 1.0 Å) indicates the correct solutions (Figure III.2.14). For SRL RNA, correct solutions with RMSD $\leq$ 1.0 Å could be determined at high resolution cut-off (2.5 - 4.0 Å), whereas the correct solutions for GIR1 RNA were determined at relatively

---

[131] E. Krissinel and K. Henrick. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **60**: 2256–2268, 2004.



(a) **(b)**

Figure III.2.13: The displaced copies of phosphorus model for (a) SRL and (b) GIR1 RNA structure shown in lattice

lower resolution range (4.0 - 5.0 Å).

The difference of the working resolution range for the SRL and GIR1 RNA (Figure III.2.14) is likely due to the difference in the anomalous strength of the two diffraction datasets. Anomalous correlation coefficients for SRL and GIR1 data are compared and showed in Figure III.2.15. High anomalous signals up to 2.0 Å resolution were found for the SRL data. The GIR1 data, in contrast, only diffracted to 2.4 Å and had lower anomalous correlations at high resolution range. Only at resolution lower than 4.0 Å, the anomalous correlations of GIR1 data were higher than 30% indicating strong enough anomalous strength in SAD sub-structure determination.[52] This is in consistent with the results shown above for the MR in anomalous differences where solutions of the GIR1 target were only found at resolution lower than 4.0 Å while solutions of the SRL target can be determined at relatively higher resolution (2.5 - 3.0 Å) (Figure III.2.14). For both SRL and DIS data, resolution higher than 5.0 Å was required to obtain solutions. With the true anomalous structures, medium to larger sized RNA sub-structure could be solved with this method.

---

[52] T. R. Schneider and G. M. Sheldrick. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **58**: 1772–1779, 2002.

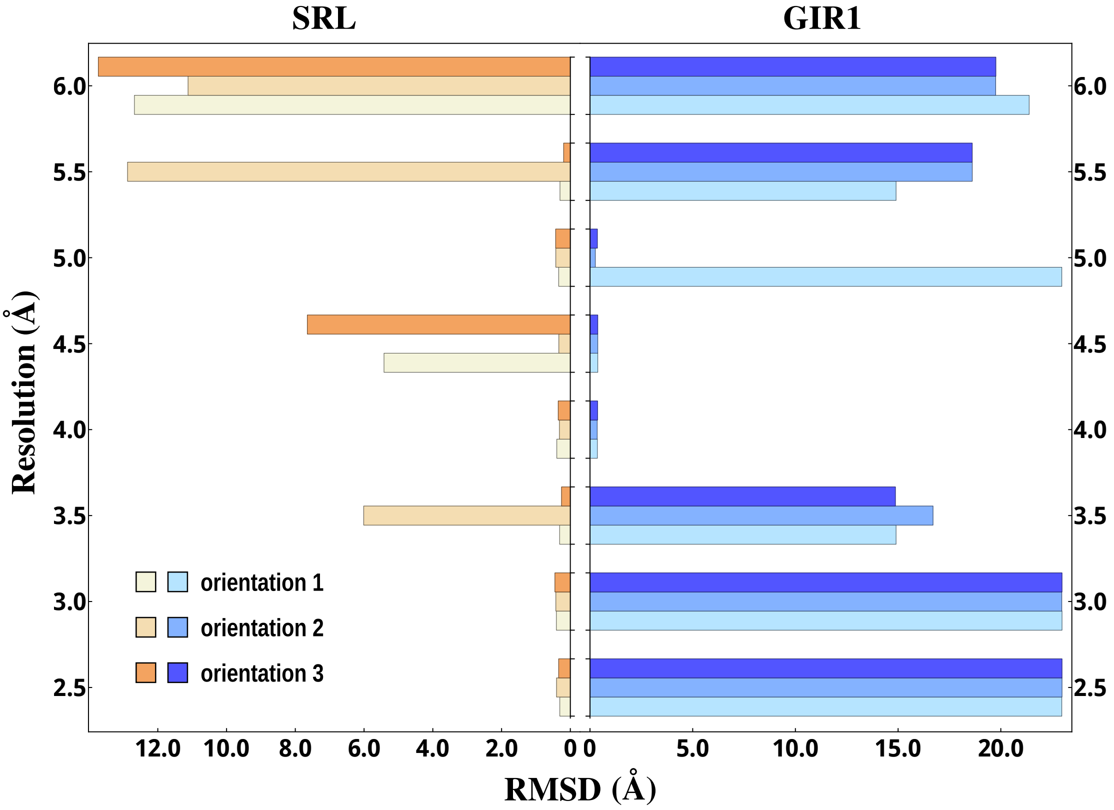


Figure III.2.14: The root-mean-square deviation (RMSD) of MR solution against the true structure for MR search with phosphorus structure in anomalous differences at different resolution cut-off from 2.5 to 6.0 Å. Left: Results of the SRL target. SRL phosphorus model was displaced in 3 different orientations. The three models were then used in MR search against anomalous differences. Right: Results of the GIR1 target. GIR1 phosphorus model was displaced in 3 distinct orientations. Each model was then tested in MR against anomalous differences. Both SRL and GIR1 data in the MR search were truncated to 2.5 to 6.0 Å with step of 0.5 Å. The RMSD between MR output coordinates and the true solution was calculated to evaluate the quality of the solutions.

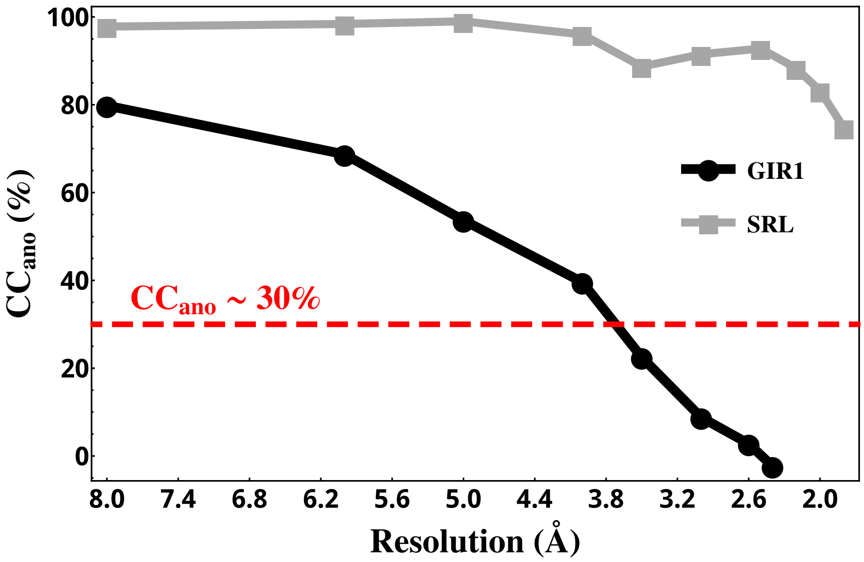


Figure III.2.15: Anomalous correlation coefficient (CC$_{ano}$) for SRL and GIR1 RNA at different resolutions. For SAD phasing in *SHELXD*, generally 25% to 30% CC$_{ano}$ is required.

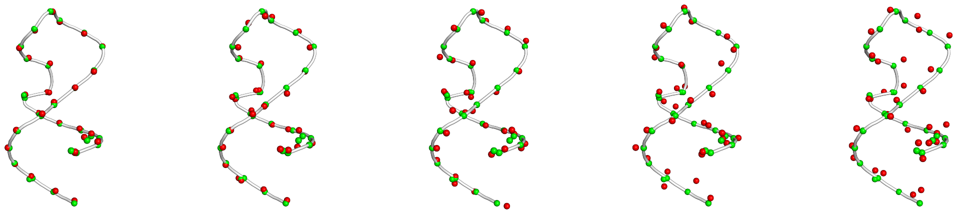
- **MR in anomalous differences with distorted P-model**

In addition to noises in the measured anomalous differences, the inaccuracy in the search model is another source of errors in MR search and strongly affects the success rate. As showed in section III.2.1.2, MR against intensities requires very accurate search model. To test the influence of the model errors on the success of MR search against anomalous differences, systematic studies were performed with SRL RNA as the search target. Randomly shifted phosphorus coordinates were used as the search model in MR. The average coordinate shifts in 3D were approximately equal to $1/\sqrt{3}$ of the root-mean-square deviation of the whole coordinates set (Table III.2.2). Mean coordinate shifts added in each model were controlled as defined levels from 0.29 to 1.44 Å which corresponded to RMSD 0.5 to 2.5 Å. Distorted SRL and GIR1 phosphors models were shown in Figure III.2.16. MR searches were performed with these models and the RMSD of the MR solutions against the correct phosphorous coordinates were calculated and showed in Table III.2.3.

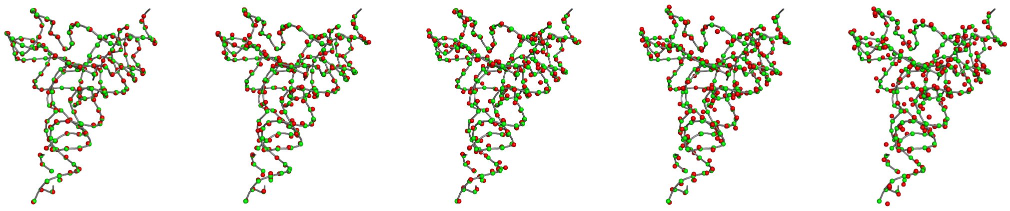
Correct MR solutions were found for SRL distorted model with 0.29 and 0.57 coordinate shifts and are marked in green in Table III.2.3a. For GIR1 distorted model, MR solution was successfully determined with a search model of 0.29 Å coordinate shifts. These values of coordinate shifts correspond to RMSD between 0.5 and 1.0 Å. As a comparison, in the case of MR search against intensities with full or partial SRL model, the MR solution could be determined with RMSD of 1.5 to 2.0 Å (Figure III.2.11). The success of MR against anomalous differences is highly sensitive to the accuracy of the sub-structure search model.

Table III.2.2: The average coordinate shifts added on the model and the corresponding root-mean-square deviation (RMSD) against true structures

σ (Å)	RMSD (Å)
0.29	0.5
0.57	1.0
0.87	1.5
1.15	2.0
1.44	2.5



(a)



(b)

Figure III.2.16: The randomly shifted phosphorus atoms (red sphere) superimposed on their original positions (green sphere). The distorted model with root-mean-square deviation (RMSD) increased from 0.5 to 2.5 Å were displaced sequentially from left to right. (a) SRL distorted phosphorus model with RMSD from 0.5 to 2.5 Å (left to right). (b) GIR1 distorted phosphorus model with RMSD from 0.5 to 2.5 Å (left to right)

Table III.2.3: The RMSD for solutions of MR search in anomalous differences with distorted phosphorus models of SRL and GIR structures. MR was performed with phosphorus model with different mean coordinate shifts. The resolutions of the data were cut from 2.5 to 5.5 Å. The RMSD values of the MR outputs against correct phosphorus coordinates were listed. Lower than 1.0 Å RMSD indicated the correct solution and were colored in green. Higher than 1.0 Å RMSD corresponded to wrong coordinates and were colored in red.

(a) SRL

Resolution (Å)	Mean coordinate shifts (Å)				
	0.29	0.57	0.87	1.15	1.44
2.5	0.37	0.28	9.7	13.1	11.4
3.5	0.24	14.2	10.2	7.42	16.7
4.5	0.31	0.33	13.4	12.5	8.21
5.5	11.2	13.9	13.9	6.75	14.6

(b) GIR1

Resolution (Å)	Mean coordinate shifts (Å)				
	0.29	0.57	0.87	1.15	1.44
4.0	0.34	14.8	13.6	12.5	17.4
4.5	0.77	20.6	8.43	12.4	13.5
5.0	15.8	9.02	18.4	12.5	12.9
5.5	16.9	6.84	17.8	12.7	7.52

III.2.3.2 MR-P-SAD with partial model

Rather than using the structure of anomalous scatters directly in MR search against anomalous differences, another approach to combine anomalous signals with MR is to use the maximum likelihood based SAD function in MR (section I.3.3). This method, called MR-SAD, has been implemented in *PHASER*.[89] MR-P-SAD refers to the MR-SAD method using anomalous signals of phosphorus.

SRL and GIR1 RNA were used as targets to test the concept. Partial model fragment in random orientation was used as input in *PHASER*. Correct partial substructure was then feed into *PHASER*. The model was then extended and completed through cycles of log-likelihood completion in *PHASER*. For SRL structure, the initial partial search model contained 6 phosphorus sites. The MR-SAD search completed all 26 sites in 2 cycles of log-likelihood completion. For GIR1 structure, the input partial model contained 124 phosphorus sites. 39 additional sites were found after 5 cycles of completion. Sites added in each cycle for these two targets were listed in Table III.2.4. Additional sites can be observed in the final electron density maps (Figure III.2.17).

This method can be considered as a basis to incorporate anomalous signals with MR. A possible strategy is to use MR-P-SAD in an iterative way as described in section III.2.2.1. With initial fragments, MR-P-SAD can be used to search the phosphorus sites in the missing part of the model. Then, the sites from the log-likelihood gain map can be manipulated and edited manually for model completion. The extended model can then be used in the next run of MR search until completion.

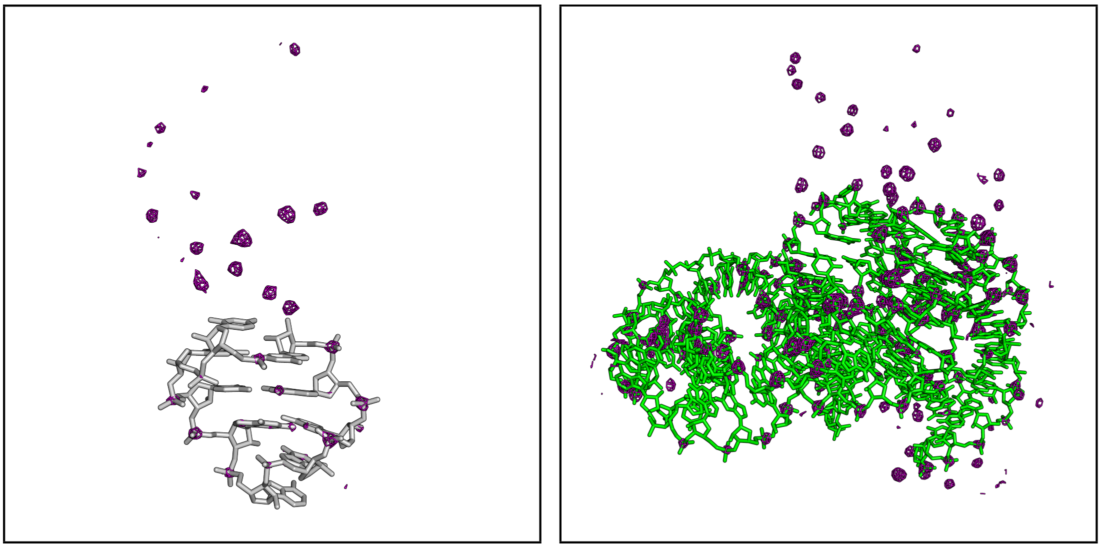
Table III.2.4: The number of sites completed in each cycle and the total number of sites after sub-structure completion with partial model MR-P-SAD in *PHASER* for (a) SRL structure and (b) GIR1 structure.

(a) SRL

SRL (26P) sub-structure completion cycle No.	Number of sites added in each cycle	Number of sites in initial model	Total number of sites after completion
1	12	6	26
2	8		

(b) GIR1

GIR1 (186P) sub-structure completion cycle No.	Number of sites added in each cycle	Number of sites in initial model	Total number of sites after completion
1	29	124	163
2	5		
3	3		
4	1		
5	1		



(a) **(b)**

Figure III.2.17: The MR-P-SAD electron density maps $(2m|F_{obs}| - D|F_{calc}|\exp(i\alpha_{calc}))$ contoured at 4.5 σ level. (a) The map obtained with SRL partial model. (b) Density map obtained with GIR1 partial model.

III.2.4 Summary and discussion

To summarize, MR with full model was successful for determining medium sized RNA structures. It was found that highly accurate model with RMSD smaller than 2.0 Å was required for the success of MR search.

The potential of applying MR search with small RNA model fragments only was tested. Partial model fragments were sufficient to solve the whole structure in the case of SRL. This approach also required extremely accurate search model.

Without homologous partial model, fragment library built from known structures could be used for MR search. The library was built with the software *ASSEMBLE*.[120] The challenge is that the correct solution can not be easily identified with the TFZ score. Iterative MR with model editing and refinement[2,119] is promising to extend the use of model fragments in MR.

Anomalous signals as an independent source of phase information can be used to support MR search. Two possible approaches were tested. With MR using phosphorous structure against anomalous differences, medium and large sized RNA (SRL and GIR1) could be successfully determined. The challenge for the generalization of this method is the requirement of high model accuracy (RMSD $<$ 1.0 Å). Test with MR-P-SAD using only partial RNA models illustrated the validity of this method.

Part IV

Conclusions and Outlook

A novel RNA structure of octameric RNA duplex (16 nucleotides) and a known structure of a sarcin-ricin loop (27 nucleotides) were successfully determined by SAD phasing using phosphorus anomalous signals (P-SAD) showing the feasibility of the method for small to medium sized nucleic acid. The requirement of P-SAD method was studied in terms of experimental parameters such as wavelength and multiplicity. Considering the weak anomalous signals of phosphorus, long wavelength would be preferable for P-SAD. However, due to the absorption effect and limitations in detector geometry/size/distance, measurable resolution is limited. 1.6 Å wavelength was chosen to compromise the gain in anomalous strength and the missing of high resolution data. High multiplicity data collection was shown to improve the data quality. However, the gain in data accuracy obtained by increasing the multiplicity may be counteracted by the effect of radiation damage with long exposure. The flux and exposure period should be adjusted to control the radiation dose in high multiplicity data collection. Multi-orientation goniometer can be used to reduce the systematic errors when multiplicity goes exceedingly high. Based on the analysis, data collection strategy for P-SAD should combine the benefit of high multiplicity, low dose, fine φ-slicing, and multi-orientation data collection with compromises between the wavelength, resolution limit and total exposure. The effectiveness of this approach has been approved by applying it to two novel RNA and protein-DNA structures.

For larger RNA structures, the applicability of the P-SAD was limited by the requirement of high resolution data in sub-structure determination. This requirement can be interpreted by considering how many observations were obtained for each phosphorus site (#Observation/#P). This is a measure of available information against the complexity of the sub-structure. About 100 - 200 unique reflections per each phosphorus atoms (#Observation/#P) were required for successful sub-structure determination. Due to the constant proportion of phosphorus in RNA molecules, #Observation/#P ratio was a constant at certain resolution for all different sized RNA. In contrast to RNA, protein-nucleic acid complex can have a lower phosphorus content, which leads to a higher #Observation/#P ratio compared to RNA molecules at the same resolution. Therefore solutions for the protein-nucleic acid complex targets can be obtained at lower resolution. This was illustrated by the success of sub-structure determination of a σ^{E2}–TGTCAAA complex (14 nucleotides and 192 amino acids) at low resolution cut-off of 4.6 Å. Considering the large amount of complex structures in Protein Data Bank,[13] P-SAD has potential to be widely applied for solving protein-nucleic acid complex structures.

When high resolution data are not available, additional approaches can supplement P-SAD in order to solve larger RNA structures. Molecular Replacement (MR) provides additional phase information and thus can be used to aid P-SAD phasing. SRL and GIR1 structures were successfully determined by MR with partial model fragments. The solutions can be then used to extract phosphorus sites for P-SAD phasing or directly used in the MR-P-SAD approach. However, highly accurate homologous models (RMSD < 2.0 Å) were required to perform the MR search. Without homologous structures, library composed of predicted model fragments can be used in MR. The challenge of this approach is to identify the solution from the incorrect outputs. With known phosphorus sites, MR can be applied in anomalous differences to determine the sub-structure for a large RNA containing 186 nucleotides. Although highly accurate model is still required, this can be a basis to extend the method in solving larger structures. It remains to be studied how to extract correct solutions from MR search using fragment library and combine it with anomalous signals towards the goal of *de novo* phasing.

Appendix

Appendix A

List of Abbreviations

ABR: Air-bearing rotation
CC: Correlation Coefficient
CC$_{ano}$: Anomalous correlation coefficient
CCD: Charge-coupled device
CMOS: Complementary metal-oxide-semiconductor
DCCM: Double Channel Cut Monochromator
DIS: Dimerization Initiation Site
DNA: Deoxyribonucleic acid
EDTA: Ethylenediaminetetraacetic acid
FMN: Flavin mononucleotide
FOM: Figure of merit
GIR1: Group I-like Ribozyme
glmS: Glucosamine-6-phosphate
HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
LLG: Log Likelihood Gain
MAD: Multi-wavelength anomalous dispersion
MIR: Multi Isomorphous Replacement
MIRAS: Single isomorphous replacement with anomalous scattering
MPD: 2-methyl-2,4-pentanediol
MR: Molecular Replacement
MW: Molecular weight
NCS: Non-crystallographic symmetry
NT: Nucleotide
PDB: Protein Data Bank
P-SAD: Phosphorus-SAD
PSI: Paul Scherrer Institut
PRIGo: Parallel Robotics Inspired Goniometer
RFN: RNA-guided FokI nuclease

RMSD: Root-mean-square deviation
RNA: Ribonucleic acid
SAD: Single-wavelength anomalous dispersion
SFOM: Site figure of merit
SIR: Single Isomorphous Replacement
SIRAS: Single isomorphous replacement with anomalous scattering
SLS: Swiss Light Source
SRL: Sarcin Ricin Loop
S-SAD: Sulfur-SAD
TFZ: Translation function Z-score
THI-box: Thiamin pyrophosphate binding riboswitch
TPP: Thiamin pyrophosphate
UTRs: Untranslated regions

Appendix B

List of Software

The programs used in this thesis are summarized in Table B.1.

Table B.1: The softwares used in data processing, structure determination and data analysis

Objective	Software
Data processing	*XDS*[106]
Sub-structure determination	*SHELXD*[52,104]
Sub-structure evaluation	*SITCOM*[110]
SAD Phasing	*PHASER*,[89] *SHELXE*[104,105]
Molecular Replacement	*PHASER*
Density modification	*DM*,[111,112]
Refinement	*PHENIX*[114]
RNA fragment Library generation	*ASSEMBLE*[120]
Data simulation	*XPREP*,[132] *SHELXPRO*[104,133]
Data analysis and plotting	*Mathematica*,[117] *BioCrystallographica*[118]

---

[106] W. Kabsch. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **66**: 125–132, 2010.

[52] T. R. Schneider and G. M. Sheldrick. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **58**: 1772–1779, 2002.

[104] G. M. Sheldrick. *Acta Crystallogr. Sect. A Found. Crystallogr.*, **64**: 112–122, 2008.

[110] F. Dall'Antonia and T. R. Schneider. *J. Appl. Crystallogr.*, **39**: 618–619, 2006.

[89] A. J. McCoy et al. *J. Appl. Crystallogr.*, **40**: 658–674, 2007.

[104] G. M. Sheldrick. *Acta Crystallogr. Sect. A Found. Crystallogr.*, **64**: 112–122, 2008.

[105] G. M. Sheldrick. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **66**: 479–485, 2010.

[111] K. Cowtan. *Jt. CCP4 ESF-EACBM Newsl. Protein Crystallogr.*, **31**: 34–38, 1994.

[112] K. Cowtan and K. Y. J. Zhang. *Prog. Biophys. Mol. Biol.*, **72**: 245–270, 1999.

[114] P. D. Adams et al. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **66**: 213–221, 2010.

[120] F. Jossinet et al. *Bioinformatics*, **26**: 2057–2059, 2010.

[132] Brucker-Nonius Inc. Madison USA *XPREP* 2004

[104] G. M. Sheldrick. *Acta Crystallogr. Sect. A Found. Crystallogr.*, **64**: 112–122, 2008.

[133] G. M. Sheldrick and T. R. Schneider. *Methods Enzymol.*, **277**: 319–343, 1997.

[117] Wolfram Research Inc. *Mathematica Edition: Version 9.0* Champaign, Illinois, 2012

[118] N. Ambert et al. *CCP4 Newsl.*, **44**: 22–31, 2006.

Appendix C

Simulated data

Simulated data were used to perform systematic analyses and compare the results to experimental data. Amplitude considering anomalous scattering for each reflection were calculated from the refined model with the software *XPREP*.[132] *SHELXPRO*[133] was used to prepare the instruction files used in *XPREP*.[132] The protocol to simulate anomalous data at specific wavelength was shown below.

1. Prepare the instruction file with *SHELXPRO*[133]
    (a) Select option [I] to initiate program setting
    (b) Define the name of the .ins file
    (c) Provide the .pdf file of refined model as input
    (d) Enter the cell parameters
    (e) Enter Z number (number of asymmetry units in the unit-cell)
    (f) Enter wavelength (Å)
    (g) Chose the default options for all the following steps

2. Calculate the amplitudes for all reflections from the refined model considering anomalous in *XPREP*[132]
    (a) Select option [G] to initiate program setting
    (b) Provide the .ins file generated by *SHELXPRO*
    (c) Define the highest resolution (Å)
    (d) Select option [F] for Firiedel opposites
    (e) Enter lattice type

---

[132] Brucker-Nonius Inc. Madison USA *XPREP* 2004

[133] G. M. Sheldrick and T. R. Schneider. *Methods Enzymol.*, **277**: 319–343, 1997.

(f) Select option [H] to enter lattice parameters
(g) Select option [S] to determine space group
(h) Select option [C] to define the chemical content of the unit-cell
(i) Select option [W] to write the output reflection file
(j) Type Enter to quit the program

Appendix D

Formulas

D.1 Structure factors and fourier transforms

$$\rho\left(\mathbf{r}\right)=\frac{1}{V}\sum F_{\mathbf{h}}e^{-2\pi(\mathbf{h}\cdot\mathbf{r})} \tag{D.1.1a}$$

$$F_{\mathbf{h}}=V\sum\rho\left(\mathbf{r}\right)e^{2\pi(\mathbf{r}\cdot\mathbf{h})} \tag{D.1.1b}$$

where, $\rho\left(\mathbf{r}\right)$ is the electron density distribution in real space and $F_{\mathbf{h}}$ is the structure factor in reciprocal space. V is the volume of the unit-cell. $\mathbf{r}$ is the real space vector and $\mathbf{h}$ is the reciprocal space vector. The electron density in real space $\rho\left(\mathbf{r}\right)$ and structure factor in reciprocal space are linked to each other via *Fourier transform*. With known structure factors in reciprocal space, one can obtain the electron density in real space via *Fourier transform*. With known electron density, the reciprocal space structure factors can be obtained via inverse *Fourier transform*.

$$F_{hkl}=|F_{hkl}|e^{i\alpha_{hkl}}=|F_{hkl}|e^{2\pi i\alpha'_{hkl}} \tag{D.1.2}$$

where h, k, l represent the miller indexes of reflections in reciprocal space and F_{hkl} is the structure factor of reflection hkl and α' is the phase angle associate to each set of coefficients.

$$\rho\left(x,y,z\right)=\frac{1}{V}\sum_{h}\sum_{k}\sum_{l}|F_{hkl}|e^{-2\pi i\left(hx+ky+lz-\alpha'_{hkl}\right)} \tag{D.1.3}$$

D.2 Direct methods

$$\varphi\left(\mathbf{h}\right)+\varphi\left(\mathbf{h}'\right)+\varphi\left(\overline{\mathbf{h}}+\overline{\mathbf{h}'}\right)\simeq 0 \tag{D.2.1a}$$

$$\tan\left[\varphi\left(\mathbf{h}\right)\right]=\frac{\sum_{h'}\kappa\left(\mathbf{h},\mathbf{h}'\right)\sin\left[\varphi\left(\mathbf{h}\right)+\varphi\left(\mathbf{h}-\mathbf{h}'\right)\right]}{\sum_{h'}\kappa\left(\mathbf{h},\mathbf{h}'\right)\cos\left[\varphi\left(\mathbf{h}\right)+\varphi\left(\mathbf{h}-\mathbf{h}'\right)\right]} \tag{D.2.1b}$$

where,

$$\kappa\left(\mathbf{h},\mathbf{h}'\right)=2\left(N^{-1/2}\right)|E\left(\mathbf{h}\right)E\left(\mathbf{h}'\right)E\left(\overline{\mathbf{h}}+\overline{\mathbf{h}'}\right)| \tag{D.2.2}$$

is the factor related to the normalized structure factors E. This relation in practice is typically used in weighted form where the contributions of the sums are weighted in terms of the probability of the relation. For non-centrosymmetric crystals, the basic phase determination procedure often starts with a selected list of large $|E|$ reflections among subgroups defined by the parity of the h, k and l indices. With a set of assigned phases, equation D.2.1a and -D.2.1b are involved to extend the phase information to other reflections with large $|E|$ values. The indication for the newly added phases is often evaluated with the variance (equation D.2.4b) of an approximate probability distribution (equation D.2.4a) predicted by D.2.1b.

$$u_{hkl}\leqslant\frac{1}{2}\left(\pm\frac{1}{2}|u_{2h,2k,2l}|\right) \tag{D.2.3}$$

The importance of this equation lies in the fact that the magnitude of u_{hkl}^2 is known and it is always positive. The only unknowns in this relation are the phases u_{2h}, u_{2k}, u_{2l}.

$$P(\varphi_{\mathbf{h}}) = \prod_{\mathbf{k}_r} P_k(\varphi_{\mathbf{h}}) \simeq A \exp\left[\sum_{\mathbf{k}_r} \kappa(\mathbf{h},\mathbf{k}) \cos(\varphi_{\mathbf{h}} - \varphi_{\mathbf{k}} - \varphi_{\mathbf{h}-\mathbf{k}})\right] \tag{D.2.4a}$$

$$V = [2\pi I_0(\alpha)]^{-1} \int_{-\pi}^{\pi} x^2 \exp(\alpha \cos x)\, \mathrm{d}x \tag{D.2.4b}$$

where,

$$\alpha = \left\{\left[\sum_{\mathbf{k}_r} \kappa(\mathbf{h},\mathbf{k}) \cos(\varphi_{\mathbf{k}} + \varphi_{\mathbf{h}-\mathbf{k}})\right]^2 + \left[\sum_{\mathbf{k}_r} \kappa(\mathbf{h},\mathbf{k}) \sin(\varphi_{\mathbf{k}} + \varphi_{\mathbf{h}-\mathbf{k}})\right]^2\right\}^{1/2}$$

and A is a normalizing constant and $\mathbf{k}_r$ represents the restricted values of k for which the corresponding normalized structure factor amplitude $|E|$ are large and I_0 is a Bessel function.[134]

D.3 Patterson method

Patterson function is a *Fourier transform* of the square of absolute values of the structure factor and in three dimensional space it can be written in vector form as,

$$\begin{aligned} P(\mathbf{u}) &= \mathscr{F}\left[|F_h|^2\right](\mathbf{u}) \\ &= \sum_h F_{\mathbf{h}}^2 e^{-2\pi i \mathbf{h}\mathbf{u}} \end{aligned} \tag{D.3.1}$$

---

[134] J. Karle and I. L. Karle. *Acta Crystallogr.*, **21**: 849–859, 1966.

With the assumption of $F_{\mathbf{h}} = F_{-\mathbf{h}}$, the expression containing only real component can be derived as,

$$P(uvw) = \frac{2}{V}\sum_h\sum_k\sum_l F_{\mathbf{h}}^2 \cos 2\pi (hu + kv + lw) \tag{D.3.2}$$

The Patterson method can be regarded as an application example of *Wiener-Khinchin Theorem*. which has wide applications in the fields of Optics, Time Series Analysis and Statistical Mechanics.[135–138] In a very general frame, the *Wiener-Khinchin Theorem* states that the autocorrelation of function $\rho(\mathbf{r})$ can be given by the *Fourier transform* of the absolute square of its *Fourier component* $F(\mathbf{h})$.[139] It can be proved as following.

A one dimensional autocorrelation function is given in the form,

$$P(u) \equiv \rho(r) \otimes \rho(-r) \equiv \int_{-\infty}^{\infty} \overline{\rho}(r)\rho(r+u)\,\mathrm{d}r \tag{D.3.3}$$

$\rho(r)$ can be represented by the *Fourier transform* of its *Fourier component* F_h and the corresponding conjugate function $\overline{\rho}(r)$ are in the forms,

$$\rho(r) = \int_{-\infty}^{\infty} F_h e^{-2\pi i h r}\,\mathrm{d}h \tag{D.3.4a}$$

$$\overline{\rho}(r) = \int_{-\infty}^{\infty} \overline{F}_h e^{2\pi i h r}\,\mathrm{d}h \tag{D.3.4b}$$

---

[135] K. Shin and P. J. Hammond. *Fundamentals of Signal Processing for Sound and Vibration Engineers*. Wiley, 2008. 244

[136] D. D. Nolte. *Optical Interferometry for Biology and Medicine: Optical Interferometry for Biology and Medicine*. Bioanalysis Springer, 2011. 104–105

[137] D. A. McQuarrie. *Statistical Mechanics*. University Science Books, 2000. 554–558

[138] A. Rogers. *Essentials of Optoelectronics with Applications*. Optical and quantum electronics series Taylor & Francis, 1997. 144

[139] E. W. Weisstein "*Wiener-Khinchin Theorem.*" *From MathWorld–A Wolfram Web Resource*. 2014

Inserting (D.3.4a) and (D.3.4b) into autocorrelation equation (D.3.3) gives,

$$\begin{aligned} P(u) &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \overline{F}_h e^{2\pi i h r}\, \mathrm{d}h \right] \left[\int_{-\infty}^{\infty} F_{h'} e^{-2\pi i h'(r+u)}\, \mathrm{d}h' \right] \mathrm{d}r \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \overline{F}_h F_{h'} e^{-2\pi i r(h'-h)} e^{-2\pi i h' u}\, \mathrm{d}r \mathrm{d}h \mathrm{d}h' \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \overline{F}_h F_{h'} \delta\left(h' - h\right) e^{-2\pi i h' u}\, \mathrm{d}h \mathrm{d}h' \\ &= \int_{-\infty}^{\infty} \overline{F}_h F_h e^{-2\pi i h r}\, \mathrm{d}h \\ &= \int_{-\infty}^{\infty} |F_h|^2 e^{-2\pi i h r}\, \mathrm{d}h \end{aligned} \tag{D.3.5}$$

Thus $\rho(r)$ relates to the square of F_h's module via a *Fourier transform*,

$$P(u) = \mathscr{F}\left[|F_h|^2\right](u) \tag{D.3.6}$$

For its application in diffraction, $P(u)$ may represent the Patterson function, which is defined as the convolution integral of electron density $\rho(r)$ over the unit cell in real space and in three dimension it is in form,

$$P(\mathbf{u}) = \int_V \rho(\mathbf{r})\, \rho(\mathbf{r} + \mathbf{u})\, \mathrm{d}\mathbf{r} \tag{D.3.7}$$

which can be written in equivalent form,

$$\begin{aligned} P(\mathbf{u}) &= \int_V \rho(\mathbf{r})\, \rho\left[\mathbf{u} - (-\mathbf{r})\right] \mathrm{d}\mathbf{r} \\ &\equiv \rho(\mathbf{r}) \otimes \rho(-\mathbf{r}) \end{aligned} \tag{D.3.8}$$

The Patterson function is shown above to be equivalent to the auto-correlation function of electron density distribution in real space and the electron density is known to be the *Fourier transform* of structure factor $F_{\mathbf{h}}$ at reciprocal space in the same form with equation D.3.4a.

D.4 Molecular Replacement

D.4.1 Patterson functions

$$\Re\left(\Omega\right) = \frac{1}{V} \int_{r_{\min}}^{r_{\max}} P_{\text{obs}}\left(\mathbf{r}\right) P_{\Omega}\left(\Omega\mathbf{r}\right) \, \mathrm{d}^3 x \tag{D.4.1.1a}$$

$$CC_E\left(\Omega\right) = \frac{\sum_{\mathbf{h}} \left(E^2_{\mathbf{h},\text{obs}} - \langle E^2_{\mathbf{h},\text{obs}} \rangle\right) \left(E^2_{\mathbf{h},\Omega} - \langle E^2_{\mathbf{h},\Omega} \rangle\right)}{\left[\sum_{\mathbf{h}} \left(E^2_{\mathbf{h},\text{obs}} - \langle E^2_{\mathbf{h},\text{obs}} \rangle\right)^2\right]^{1/2} \left[\sum_{\mathbf{h}} \left(E^2_{\mathbf{h},\Omega} - \langle E^2_{\mathbf{h},\Omega} \rangle\right)^2\right]^{1/2}} \tag{D.4.1.1b}$$

P_{obs} and P_{Ω} are the observed Patterson and Patterson of search model at orientation Ω. E_{obs} and E_{Ω} are the normalized structure factors calculated from observed data and model. It is worth to note that, E_{Ω} in the Paterson correlation (equation D.4.1.1b) or so called direct Paterson rotation function, are calculated from the search model and therefore depends on both intra- and inter-molecular distances. Therefore the rotation search using Patterson correlation is not always more accurate than other rotation functions.[140]

$$T\left(\mathbf{t}\right) = \int_V \left[P_{\text{obs}}\left(\mathbf{u}\right) - \sum_{j=1}^{N_{\text{sym}}} P_{jj}\left(u\right)\right] \left[P_T\left(\mathbf{u},\mathbf{t}\right) - \sum_{j=1}^{N_{\text{sym}}} P_{jj}\left(\mathbf{u}\right)\right] \mathrm{d}\mathbf{u} \tag{D.4.1.2}$$

where P_{jj} refers to Patterson functions calculated with self-vectors only.

---

[140] B. Rupp. *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology*. Garland Science, 2010. 556–557

D.4.2 Maximum-likelihood functions

$$\begin{aligned}\mathbf{F}(\mathbf{h}) &= \sum_j \sum i \exp\left[2\pi i \mathbf{h}\cdot\left(\mathbf{C}_j \mathbf{x}_i + \mathbf{d}_j\right)\right] \\ &= \sum_j \exp\left(2\pi i \mathbf{h}\cdot\mathbf{d}_j\right)\mathbf{F}(\mathbf{h}, j)\end{aligned} \tag{D.4.2.1}$$

where C_j and d_j are the rotation and translation components of the j^{th} symmetry operation, x_j are the fractional coordinates and $\mathbf{F}(h, j)$ is the transform of the molecule in complex form corresponding to the j^{th} symmetry operation.[141]

$$P_a\left(\mathbf{F}_o|\mathbf{F}_c\right) = \frac{1}{\pi\varepsilon\sigma_\Delta^2}\exp\left[-\frac{|\mathbf{F}_o - D\mathbf{F}_c|^2}{\varepsilon\sigma_\Delta^2}\right] \tag{D.4.2.2a}$$

$$P_c\left(\mathbf{F}_o|\mathbf{F}_c\right) = \frac{1}{\left(2\pi\varepsilon\sigma_\Delta^2\right)^{1/2}}\exp\left[-\frac{|\mathbf{F}_o - D\mathbf{F}_c|^2}{2\varepsilon\sigma_\Delta^2}\right] \tag{D.4.2.2b}$$

where σ_Δ^2 is the weighting factor and ε is the expected intensity factor calculated from Luzzati weighting.[142] The equation D.4.2.2a is the distribution of acentric structure factors and equation D.4.2.2b is for centric structure factors. As the phases of experimental data are not known, the phases component in the likelihood functions should be integrated out during the calculation and the two factors σ_Δ and D related to Luzzati weighting can be summarized with a single parameter σ_A and gives,

$$P_a\left(E_o|E_c\right) = \frac{2E_o}{1-\sigma_A^2}\exp\left(-\frac{E_o^2 + \sigma_A^2 E_C^2}{1-\sigma_A^2}\right) I_0\left(\frac{2E_o\sigma_A E_c}{1-\sigma_A^2}\right) \tag{D.4.2.3a}$$

$$P_a\left(E_o|E_c\right) = \left[\frac{2}{\pi\left(1-\sigma_A^2\right)}\right]^{1/2}\exp\left[-\frac{E_o^2 + \sigma_A^2 E_C^2}{2\left(1-\sigma_A^2\right)}\right]\cosh\left(\frac{E_o\sigma_A E_c}{1-\sigma_A^2}\right) \tag{D.4.2.3b}$$

where $\mathbf{E}$ are related to Luzzati weighting factor with $\mathbf{E} = \mathbf{F}_o/(\varepsilon\Sigma_N)^{1/2}$ and $\mathbf{E} =$

---

[141] P. Evans and A. McCoy. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **64**: 1–10, 2008.

[142] V. Luzzati. *Acta Crystallogr.*, **5**: 802–810, 1952.

$\mathbf{F}_o/(\varepsilon\Sigma_P)^{1/2}$. $E_o = |\mathbf{E}_o|$ and $E_c = |\mathbf{E}_c|$ and $\sigma_A = D\,(\Sigma_P/\Sigma_N)^{1/2}$.

D.5 Data quality parameters

$$R_{\text{merge}} = \sum_{\mathbf{h}}\sum_{l}|I_{hl} - \langle I_{\mathbf{h}}\rangle| / \sum_{\mathbf{h}}\sum_{l}\langle I_{\mathbf{h}}\rangle \tag{D.5.1}$$

$$R_{\text{meas}} = \sum_{\mathbf{h}}\left(\frac{n_h}{n_h - 1}\right)\sum_{l}|I_{hl} - \langle I_{\mathbf{h}}\rangle| / \sum_{\mathbf{h}}\sum_{l}\langle I_{\mathbf{h}}\rangle \tag{D.5.2}$$

where $\langle I_{\mathbf{h}}\rangle$ is the average intensity of reflection **h**. The I_{hl} is the intensity of the l measurement of reflection **h**. n_h is the number of observations of reflection **h**. Both R_{merge} and R_{meas} are the measures of how well the different observations agree. The R_{merge} is not a particularly good measure of data quality as it only measures the discrepancy between observations and takes no account of the improvement in the merged intensity by averaging many observations.[143] Therefore R_{merge} tends to increase with increasing multiplicity. R_{meas} is an improved version of R_{merge} which is independent to the data multiplicity.[125]

$$\Delta F_{\text{ano}}/F = 2^{1/2}\left(f'' N_A^{1/2}\right) / \left(6.7 N_P^{1/2}\right) \tag{D.5.3}$$

where ΔF_{ano} is the anomalous difference and F is the amplitude of the reflections. f'' is the theoretical value of the imaginary component for the atomic scattering factors of phosphorus. N_A is the number of the anomalous scatters and N_P is the number of the other atoms within the whole molecule. $\Delta F_{\text{ano}}/F$ is used to

---

[143] P. Evans. *Acta Crystallogr. D. Biol. Crystallogr.*, **62**: 72–82, 2006.

[125] K. Diederichs and P. A. Karplus. *Nature*, **4**: 269–275, 1997.

estimate the anomalous strength of the overall structure.

$$\mathrm{CC} = \frac{\sum wE_oE_c\sum w - \sum wE_o\sum wE_c}{\sqrt{\left[\sum wE_o^2\sum w - \left(\sum wE_o\right)^2\right]\left[\sum wE_c^2\sum w - \left(\sum wE_c^2\right)^2\right]}} \tag{D.5.4}$$

where E_o is the amplitude of the normalized observed structure factor and E_c is the amplitude of the normalized calculated structure factor. w is the weighting factor to weight down influence of the reflections with the less reliable anomalous differences. In *SHELXD* w is set to $\left[1+g\sigma^2(E)\right]^{-1}$, with a default value of 0.1 for g.[52] CC is the correlation coefficient between E_o and E_c. High value of CC identifies potential sub-structure solutions in *SHELXD*.

D.6 FOM (from SITCOM)

$$\mathrm{SFOM} = \frac{1}{w_\mathrm{d}+w_\mathrm{pk}+w_\mathrm{cr}}\left(w_\mathrm{d}\tilde{d}_\mathrm{mean}+w_\mathrm{pk}h_\mathrm{pk}+w_\mathrm{cr}\mathrm{cr}\right) \tag{D.6.1}$$

where $\tilde{d} = 1/(1+d_\mathrm{mean})$, with d_mean being the mean distance between equivalent sites; h_pk was the average peak height for the unique sites; cr was the nubmer of hits divided by the number of total input sites; The defaults values of the weighting factors were $w_\mathrm{d} = 0.3$, $w_\mathrm{pk} = 0$, and $w_\mathrm{cr} = 1.0$.

---

[52] T. R. Schneider and G. M. Sheldrick. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **58**: 1772–1779, 2002.

Appendix E

List of Figures

Appendix F

List of Tables

Bibliography

[1] Z. Dauter and D. A. Adamiak. Anomalous signal of phosphorus used for phasing DNA oligomer: importance of data redundancy. *Acta Crystallographica Section D: Biological Crystallography*, **57**: 990–995, 2001. doi: 10.1107/S0907444901006382 (see pp. 1, 3, 33, 78, 80)

[2] M. P. Robertson and W. G. Scott. A general method for phasing novel complex RNA crystal structures without heavy-atom derivatives. *Acta Crystallographica Section D: Biological Crystallography*, **D64**: 738–744, 2008. doi: 10.1107/S0907444908011578 (see pp. 1, 3, 24, 62, 64, 112, 123)

[3] F. H. C. Crick. On protein synthesis. *Symposia of the Society for Experimental Biology*, **XII**: 139–163, 1956. (see p. 7)

[4] F. H. C. Crick. Central Dogma of Molecular Biology. *Nature*, **227**: 561–563, 1970. doi: 10.1038/227561a0 (see p. 7)

[5] J. D. Watson and F. H. C. Crick. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*, **171**: 737–738, 1953. doi: 10.1038/171737a0 (see pp. 9, 13)

[6] K. Hoogsteen. The crystal and molecular structure of a hydrogen-bonded complex between 1-methylthymine and 9-methyladenine. *Acta Crystallographica*, **16**: 907–916, 1963. doi: 10.1107/S0365110X63002437 (see p. 10)

[7] D. K. Hendrix, S. E. Brenner, and S. R. Holbrook. RNA structural motifs: building blocks of a modular biomolecule. *Quarterly Reviews of Biophysicsiophysics*, **38**: 221–243, 2005. doi: 10.1017/S0033583506004215 (see pp. 10–12)

[8] P. B. Moore. Structural motifs in RNA. *Annual Review of Biochemistry*, **68**: 287–300, 1999. doi: 10.1146/annurev.biochem.68.1.287 (see p. 11)

[9] C. Woese, S. Winker, and R. Gutell. Architecture of ribosomal RNA: constraints on the sequence of "tetra-loops". *Proceedings of the National Academy of Sciences of the United States of America*, **87**: 8467–8471, 1990. doi: 10.1073/pnas.87.21.8467 (see p. 11)

[10] A. Liljas. *Textbook Of Structural Biology*. World Scientific Publishing Company, Incorporated, 2009. 78–88 (see pp. 11, 12)

[11] D. J. Klein, T. M. Schmeing, P. B. Moore, and T. A. Steitz. The kink-turn: a new RNA secondary structure motif. *The EMBO Journal*, **20**: 4214–4221, 2001. doi: 10.1093/emboj/20.15.4214 (see p. 12)

[12] K. T. Schroeder, P. Daldrop, and D. M. J. Lilley. RNA tertiary interactions in a riboswitch stabilize the structure of a kink turn. *Structure*, **19**: 1233–1240, 2011. doi: 10.1016/j.str.2011.07.003 (see pp. 12, 13)

[13] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. The Protein Data Bank. *Nucleic Acids Research*, **28**: 235–242, 2000. doi: 10.1093/nar/28.1.235 (see pp. 13, 18, 62, 66, 112, 128)

[14] R. W. Holley, J. Apgar, G. A. Everett, J. T. Madison, M. Marquisee, S. H. Merrill, J. R. Penswick, and A. Zamir. Structure of a Ribonucleic Acid. *Science*, **147**: 1462–1465, 1965. doi: 10.1126/science.147.3664.1462 (see p. 14)

[15] S. H. Kim, G. Quigley, F. L. Suddath, and A. Rich. High-resolution x-ray diffraction patterns of crystalline transfer RNA that show helical regions. *Proceedings of the National Academy of Sciences of the United States of America*, **68**: 841–845, 1971. doi: 10.1073/pnas.68.4.841 (see p. 14)

[16] S. H. Kim, G. J. Quigley, F. L. Suddath, A. McPherson, D. Sneden, J. J. Kim, J. Weinzierl, and A. Rich. Three-dimensional structure of yeast phenylalanine transfer RNA: folding of the polynucleotide chain. *Science*, **179**: 285–288, 1973. doi: 10.1126/science.179.4070.285 (see p. 14)

[17] N. A. Campbell and J. B. Reece. *Biology*. Pearson, Benjamin Cummings, 2005. 360–365 (see p. 15)

[18] S. Tarakhovsky *Backstage with a command performer* 2003 (see p. 15)

[19] B. Alberts, A. Johnson, J. Lewis, P. Walter, M. Raff, and K. Roberts. *Molecular Biology of the Cell 4th Edition: International Student Edition*. Routledge, 2002. 207 (see p. 15)

[20] T. J. Richmond, J. T. Finch, B. Rushton, D. Rhodes, and A. Klug. Structure of the nucleosome core particle at 7 Å resolution. *Nature*, **311**: 532–537, 1984. doi: 10.1038/311532a0 (see p. 15)

[21] K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, and T. J. Richmond. Crystal structure of the nucleosome core particle at 2.8

Å resolution. *Nature*, **389**: 251–260, 1997. doi: 10.1038/38444 (see p. 15)

[22] E. L. Palmer, A. Gewiess, J. M. Harp, M. H. York, and G. J. Bunick. Large-scale production of palindrome DNA fragments. *Analytical Biochemistry*, **231**: 109–114, 1995. doi: 10.1006/abio.1995.1509 (see p. 15)

[23] J. M. Harp, E. C. Uberbacher, A. E. Roberson, E. L. Palmer, A. Gewiess, and G. J. Bunick. X-ray diffraction analysis of crystals containing twofold symmetric nucleosome core particles. *Acta Crystallographica Section D: Biological Crystallography*, **52**: 283–288, 1996. doi: 10.1107/S0907444995009139 (see p. 15)

[24] C. L. White, R. K. Suto, and K. Luger. Structure of the yeast nucleosome core particle reveals fundamental changes in internucleosome interactions. *The EMBO Journal*, **20**: 5207–5218, 2001. doi: 10.1093/emboj/20.18.5207 (see p. 15)

[25] K. Kruger, P. J. Grabowski, A. J. Zaug, J. Sands, D. E. Gottschling, and T. R. Cech. Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. *Cell*, **31**: 147–157, 1982. doi: 10.1016/0092-8674(82)90414-7 (see p. 15)

[26] M. Martick, L. H. Horan, H. F. Noller, and W. G. Scott. A discontinuous hammerhead ribozyme embedded in a mammalian messenger RNA. *Nature*, **454**: 899–902, 2008. doi: 10.1038/nature07117 (see p. 15)

[27] K. S. Keating, N. Toor, P. S. Perlman, and A. M. Pyle. A structural analysis of the group II intron active site and implications for the spliceosome. *RNA*, **16**: 1–9, 2010. doi: 10.1261/rna.1791310 (see p. 16)

[28] M. P. Robertson and W. G. Scott. The structural basis of ribozyme-catalyzed RNA assembly. *Science*, **315**: 1549–1553, 2007. doi: 10.1126/science.1136231 (see p. 16)

[29] E. Nudler and A. S. Mironov. The riboswitch control of bacterial metabolism. *Trends in Biochemical Sciences*, **29**: 11–17, 2004. doi: 10.1016/j.tibs.2003.11.004 (see p. 16)

[30] A. G. Vitreschak, D. A. Rodionov, A. A. Mironov, and M. S. Gelfand. Riboswitches: the oldest mechanism for the regulation of gene expression? *Trends in Genetics*, **20**: 44–50, 2004. doi: 10.1016/j.tig.2003.11.008 (see p. 16)

[31] R. T. BATEY. Structures of regulatory elements in mRNAs. *Current Opinion in Structural Biology*, **16**: 299–306, 2006. doi: 10.1016/j.sbi.2006.05.001 (see p. 16)

[32] B. J. TUCKER and R. R. BREAKER. Riboswitches as versatile gene control elements. *Current Opinion in Structural Biology*, **15**: 342–348, 2005. doi: 10.1016/j.sbi.2005.05.003 (see p. 16)

[33] A. S. MIRONOV, I. GUSAROV, R. RAFIKOV, L. E. LOPEZ, K. SHATALIN, R. A. KRENEVA, D. A. PERUMOV, and E. NUDLER. Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. *Cell*, **111**: 747–756, 2002. doi: 10.1016/S0092-8674(02)01134-0 (see p. 16)

[34] A. NAHVI, N. SUDARSAN, M. S. EBERT, X. ZOU, K. L. BROWN, and R. R. BREAKER. Genetic control by a metabolite binding mRNA. *Chemistry & Biology*, **9**: 1043–1049, 2002. doi: 10.1016/S1074-5521(02)00224-7 (see p. 16)

[35] W. C. WINKLER, S. COHEN-CHALAMISH, and R. R. BREAKER. An mRNA structure that controls gene expression by binding FMN. *Proceedings of the National Academy of Sciences of the United States of America*, **99**: 15908–15913, 2002. doi: 10.1073/pnas.212628899 (see p. 16)

[36] W. WINKLER, A. NAHVI, and R. R. BREAKER. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature*, **419**: 952–956, 2002. doi: 10.1038/nature01145 (see p. 16)

[37] F. J. GRUNDY and T. M. HENKIN. The S box regulon: a new global transcription termination control system for methionine and cysteine biosynthesis genes in gram-positive bacteria. *Molecular Microbiology*, **30**: 737–749, 1998. doi: 10.1046/j.1365-2958.1998.01105.x (see p. 16)

[38] J. MIRANDA-RÍOS, M. NAVARRO, and M. SOBERÓN. A conserved RNA structure (thi box) is involved in regulation of thiamin biosynthetic gene expression in bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, **98**: 9736–9741, 2001. doi: 10.1073/pnas.161168098 (see p. 17)

[39] M. S. GELFAND, A. A. MIRONOV, J. JOMANTAS, Y. I. KOZLOV, and D. A. PERUMOV. A conserved RNA structure element involved in the regulation of bacterial riboflavin synthesis genes. *Trends in Genetics*, **15**: 439–442, 1999. doi: 10.1016/S0168-9525(99)01856-9 (see p. 17)

[40] C. V. FRANKLUND and R. J. KADNER. Multiple transcribed elements control expression of the Escherichia coli btuB gene. *Journal of Bacteriology*, **179**: 4039–4042, 1997. (see p. 17)

[41] J. E. Barrick, K. A. Corbino, W. C. Winkler, A. Nahvi, M. Mandal, J. Collins, M. Lee, A. Roth, N. Sudarsan, I. Jona, J. K. Wickiser, and R. R. Breaker. New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proceedings of the National Academy of Sciences of the United States of America*, **101**: 6421–6426, 2004. doi: 10.1073/pnas.0308014101 (see p. 17)

[42] Z. Weinberg, J. X. Wang, J. Bogue, J. Yang, K. Corbino, R. H. Moy, and R. R. Breaker. Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. *Genome Biology*, **11**: R31, 2010. doi: 10.1186/gb-2010-11-3-r31 (see p. 17)

[43] Z. Weinberg, J. E. Barrick, Z. Yao, A. Roth, J. N. Kim, J. Gore, J. X. Wang, E. R. Lee, K. F. Block, N. Sudarsan, S. Neph, M. Tompa, W. L. Ruzzo, and R. R. Breaker. Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. *Nucleic Acids Research*, **35**: 4809–4819, 2007. doi: 10.1093/nar/gkm487 (see p. 17)

[44] F. Schluenzen, A. Tocilj, R. Zarivach, J. Harms, M. Gluehmann, D. Janell, A. Bashan, H. Bartels, I. Agmon, F. Franceschi, and A. Yonath. Structure of functionally activated small ribosomal subunit at 3.3 Å resolution. *Cell*, **102**: 615–623, 2000. doi: 10.1016/S0092-8674(00)00084-2 (see p. 17)

[45] N. Ban, P. Nissen, J. Hansen, P. B. Moore, and T. A. Steitz. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*, **289**: 905–920, 2000. doi: 10.1126/science.289.5481.905 (see p. 17)

[46] B. T. Wimberly, D. E. Brodersen, W. M. Clemons Jr, R. J. Morgan-Warren, A. P. Carter, C. Vonrhein, T. Hartsch, and V. Ramakrishnan. Structure of the 30S ribosomal subunit. *Nature*, **407**: 327–339, 2000. doi: 10.1038/35030006 (see p. 17)

[47] N. E. Chayen and E. Saridakis. Protein crystallization: from purified protein to diffraction-quality crystal. *Nature Methods*, **5**: 147–153, 2008. doi: 10.1038/NMETH.F.203 (see pp. 18, 19)

[48] J. Karle and H. Hauptman. The phases and magnitudes of the structure factors. *Acta Crystallographica*, **3**: 181–187, 1950. doi: 10.1107/S0365110X50000446 (see p. 22)

[49] R. J. Morris and G. Bricogne. Sheldrick's 1.2 Å rule and beyond. *Acta Crystallographica Section D: Biological Crystallography*, **59**: 615–617, 2003. doi: 10.1107/S090744490300163X (see p. 22)

[50] G. M. SHELDRICK. Phase annealing in SHELX-90: direct methods for larger structures. *Acta Crystallographica Section A: Foundations of Crystallography*, **46**: 467–473, 1990. doi: 10.1107/S0108767390000277 (see p. 22)

[51] R. MILLER, S. M. GALLO, H. G. KHALAK, and C. M. WEEKS. SnB: crystal structure determination via shake-and-bake. *Journal of Applied Crystallography*, **27**: 613–621, 1994. doi: 10.1107/S0021889894000191 (see p. 22)

[52] T. R. SCHNEIDER and G. M. SHELDRICK. Substructure solution with SHELXD. *Acta Crystallographica Section D: Biological Crystallography*, **58**: 1772–1779, 2002. doi: 10.1107/S0907444902011678 (see pp. 22, 58, 74, 83, 88, 116, 133, 145)

[53] E. D. L. FORTELLE and G. BRICOGNE. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods in Enzymology*, **276**: 472–494, 1997. doi: 10.1016/S0076-6879(97)76073-7 (see p. 22)

[54] A. VAGIN and A. TEPLYAKOV. MOLREP: an Automated Program for Molecular Replacement. *Journal of Applied Crystallography*, **30**: 1022–1025, 1997. doi: 10.1107/S0021889897006766 (see p. 22)

[55] R. W. GROSSE-KUNSTLEVE and P. D. ADAMS. Substructure search procedures for macromolecular structures. *Acta Crystallographica Section D: Biological Crystallography*, **59**: 1966–1973, 2003. doi: 10.1107/S0907444903018043 (see p. 22)

[56] M. G. ROSSMANN and D. M. BLOW. The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallographica*, **15**: 24–31, 1962. doi: 10.1107/S0365110X62000067 (see p. 23)

[57] A. L. PATTERSON. A direct method for the determination of the components of interatomic distances in crystals. *Zeitschrift für Kristallographie-Crystalline Materials*, **90**: 517–542, 1935. doi: 10.1524/zkri.1935.90.1.517 (see p. 23)

[58] V. Y. LUNIN and A. G. URZHUMTSEV. Improvement of protein phases by coarse model modification. *Acta Crystallographica Section A: Foundations of Crystallography*, **40**: 269–277, 1984. doi: 10.1107/S0108767384000544 (see p. 23)

[59] R. J. READ. Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallographica Section A: Foundations of Crystallography*, **42**: 140–149, 1986. doi: 10.1107/S0108767386099622 (see p. 23)

[60] I. BAIKALOV and R. E. DICKERSON. Molecular Replacement Using DNA Helical Symmetry. *Acta Crystallographica Section D: Biological Crystallography*, **54**: 324–333, 1998. doi: 10.1107/S0907444997010512 (see p. 24)

[61] E. DODSON. The befores and afters of molecular replacement. *Acta crystallographica Section D: Biological Crystallography*, **64**: 17–24, 2008. doi: 10.1107/S0907444907049736 (see p. 24)

[62] D. W. GREEN, V. M. INGRAM, and M. F. PERUTZ. The structure of haemoglobin. IV. Sign determination by the isomorphous replacement method. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **225**: 287–307, 1954. doi: 10.1098/rspa.1954.0203 (see p. 24)

[63] D. HARKER. The determination of the phases of the structure factors of non-centrosymmetric crystals by the method of double isomorphous replacement. *Acta Crystallographica*, **9**: 1–9, 1956. doi: 10.1107/S0365110X56000012 (see p. 25)

[64] M. VIJAYAN and S. RAMASESHAN. Anomalous-scattering method. *International Tables for Crystallography*, **B**: 265–269, 2006. doi: 10.1107/97809553602060000557 (see p. 26)

[65] G. FRIEDEL. Sur les symétries cristallines que peut révéler la diffraction des rayons Röntgen. *Comptes Rendus de l'Académie des Sciences*, **157**: 1533–1536, 1913. (see p. 26)

[66] D. CARVIN, S. A. ISLAM, M. J. E. STERNBERG, and T. L. BLUNDELL. ISOMORPHOUS REPLACEMENT - The preparation of heavy-atom derivatives of protein crystals for use in multiple isomorphous replacement and anomalous scattering. *International Tables for Crystallography*, **F**: 247–255, 2006. doi: 10.1107/97809553602060000679 (see pp. 29, 30)

[67] F. E. REYES, A. D. GARST, and R. T. BATEY. Strategies in RNA crystallography. *Methods in Enzymology*, **469**: 119–139, 2009. doi: 10.1016/S0076-6879(09)69006-6 (see p. 29)

[68] J. H. CATE and J. A. DOUDNA. Solving large RNA structures by X-ray crystallography. *Methods in Enzymology*, **317**: 169–80, 2000. doi: 10.1016/S0076-6879(00)17014-4 (see p. 29)

[69] A. Y. KEEL, R. P. RAMBO, R. T. BATEY, and J. S. KIEFT. A general strategy to solve the phase problem in RNA crystallography. *Structure*, **15**: 761–772, 2007. doi: 10.1016/j.str.2007.06.003 (see p. 29)

[70] J. KARLE. Some developments in anomalous dispersion for the structural investigation of macromolecular systems in biology. *International*

Journal of Quantum Chemistry, **18**: 357–367, 1980. doi: 10.1002/qua.560180734 (see p. 31)

[71] W. A. HENDRICKSON, A. PÄHLER, J. L. SMITH, Y. SATOW, E. A. MERRITT, and R. P. PHIZACKERLEY. Crystal structure of core streptavidin determined from multiwavelength anomalous diffraction of synchrotron radiation. *Proceedings of the National Academy of Sciences of the United States of America*, **86**: 2190–2194, 1989. doi: 10.1073/pnas.86.7.2190 (see p. 31)

[72] W. A. HENDRICKSON, J. R. HORTON, and D. M. LEMASTER. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *The EMBO Journal*, **9**: 1665–1672, 1990. (see p. 31)

[73] W. A. HENDRICKSON and M. M. TEETER. Structure of the hydrophobic protein crambin determined directly from the anomalous scattering of sulphur. *Nature*, **290**: 107–113, 1981. doi: 10.1038/290107a0 (see p. 32)

[74] Q. LIU, T. DAHMANE, Z. ZHANG, Z. ASSUR, J. BRASCH, L. SHAPIRO, F. MANCIA, and W. A. HENDRICKSON. Structures from anomalous diffraction of native biological macromolecules. *Science*, **336**: 1033–1037, 2012. doi: 10.1126/science.1218753 (see pp. 32, 35)

[75] Q. LIU, Q. LIU, and W. A. HENDRICKSON. Robust structural analysis of native biological macromolecules from multi-crystal anomalous diffraction data. *Acta Crystallographica Section D: Biological Crystallography*, **69**: 1314–1332, 2013. doi: 10.1107/S0907444913001479 (see p. 32)

[76] E. MICOSSI, W. N. HUNTER, and G. A. LEONARD. De novo phasing of two crystal forms of tryparedoxin II using the anomalous scattering from S atoms: a combination of small signal and medium resolution reveals this to be a general tool for solving protein crystal structures. *Acta Crystallographica Section D: Biological Crystallography*, **58**: 21–28, 2001. doi: 10.1107/S0907444901016808 (see p. 33)

[77] B. C. WANG. Resolution of phase ambiguity in macromolecular crystallography. *Methods in Enzymology*, **115**: 90–112, 1985. doi: 10.1016/0076-6879(85)15009-3 (see p. 33)

[78] D. T. CROMER. Anomalous dispersion corrections computed from self-consistent field relativistic Dirac-Slater wave functions. *Acta Crystallographica*, **18**: 17–23, 1965. doi: 10.1107/S0365110X6500004X (see p. 34)

[79] D. T. CROMER and J. B. MANN. X-ray scattering factors computed from numerical Hartree-Fock wave functions. *Acta Crystallographica Section A: Foundations of Crystallography*, **24**: 321–324, 1968. doi: 10.1107/S0567739468000550 (see p. 34)

[80] D. T. CROMER and D. A. LIBERMAN. Relativistic calculation of anomalous scattering factors for X rays. *The Journal of Chemical Physics*, **53**: 1891–1898, 1970. doi: 10.1063/1.1674266 (see p. 34)

[81] D. T. CROMER and D. A. LIBERMAN. Anomalous dispersion calculations near to and on the long-wavelength side of an absorption edge. *Acta Crystallographica Section A: Foundations of Crystallography*, **37**: 267–268, 1981. doi: 10.1107/S0567739481000600 (see p. 34)

[82] V. MYKHAYLYK and A. WAGNER. Towards long-wavelength protein crystallography: keeping a protein crystal frozen in vacuum. *Journal of Physics: Conference Series*, **425**: 012010, 2013. doi: 10.1088/1742-6596/425/1/012010 (see p. 34)

[83] R. SANISHVILI, D. W. YODER, S. B. POTHINENI, G. ROSENBAUM, S. XU, S. VOGT, S. STEPANOV, O. A. MAKAROV, S. CORCORAN, R. BENN, V. NAGARAJAN, J. L. SMITH, and R. F. FISCHETTI. Radiation damage in protein crystals is reduced with a micron-sized X-ray beam. *Proceedings of the National Academy of Sciences of the United States of America*, **108**: 6127–6132, 2011. doi: 10.1073/pnas.1017701108 (see p. 35)

[84] K. DJINOVIĆ CARUGO, J. R. HELLIWELL, H. STUHRMANN, and M. S. WEISS. Softer and soft X-rays in macromolecular crystallography. *Journal of Synchrotron Radiation*, **12**: 410–419, 2005. doi: 10.1107/S0909049504025762 (see p. 35)

[85] J. P. SCHUERMANN and J. J. TANNER. MRSAD: using anomalous dispersion from S atoms collected at Cu Kalpha wavelength in molecular-replacement structure determination. *Acta Crystallographica Section D: Biological Crystallography*, **59**: 1731–1736, 2003. doi: 10.1107/S0907444903015725 (see p. 37)

[86] E. N. BAKER, B. F. ANDERSON, A. J. DOBBS, and E. J. DODSON. Use of iron anomalous scattering with multiple models and data sets to identify and refine a weak molecular replacement solution: structure analysis of cytochrome c’ from two bacterial species. *Acta crystallographica Section D: Biological Crystallography*, **51**: 282–289, 1995. doi: 10.1107/S0907444994012874 (see p. 37)

[87] J. UNGE, C. MUELLER-DIECKMANN, S. PANJIKAR, P. A. TUCKER, V. S. LAMZIN, and M. S. WEISS. On the routine use of soft X-rays in macromolecular crystallography. Part V. Molecular replacement and anoma-

lous scattering. *Acta crystallographica Section D: Biological Crystallography*, **67**: 729–738, 2011. doi: 10.1107/S0907444911024887 (see p. 37)

[88] A. J. McCoy, L. C. Storoni, and R. J. Read. Simple algorithm for a maximum-likelihood SAD function. *Acta crystallographica Section D: Biological Crystallography*, **60**: 1220–1228, 2004. doi: 10.1107/S0907444904009990 (see p. 37)

[89] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, and R. J. Read. Phaser crystallographic software. *Journal of Applied Crystallography*, **40**: 658–674, 2007. doi: 10.1107/S0021889807021206 (see pp. 37, 59, 61, 76, 100, 103, 115, 121, 133)

[90] V. Olieric, U. Rieder, K. Lang, A. Serganov, C. Schulze-Briese, R. Micura, P. Dumas, and E. Ennifar. A fast selenium derivatization strategy for crystallization and phasing of RNA structures. *RNA*, 707–715, 2009. doi: 10.1261/rna.1499309 (see pp. 43, 45)

[91] C. C. Correll, I. G. Wool, and A. Munishkin. The two faces of the Escherichia coli 23 S rRNA sarcin/ricin domain: the structure at 1.11 Å resolution. *Journal of Molecular Biology*, **292**: 275–287, 1999. doi: 10.1006/jmbi.1999.3072 (see pp. 43, 45)

[92] E. Ennifar, M. Yusupov, P. Walter, R. Marquet, B. Ehresmann, C. Ehresmann, and P. Dumas. The crystal structure of the dimerization initiation site of genomic HIV-1 RNA reveals an extended duplex with two adenine bulges. *Structure*, **7**: 1439–1449, 1999. doi: 10.1016/S0969-2126(00)80033-7 (see pp. 43, 45, 66)

[93] M. Meyer, H. Nielsen, V. Oliéric, P. Roblin, S. D. Johansen, E. Westhof, and B. Masquida. Speciation of a group I intron into a lariat capping ribozyme. *Proceedings of the National Academy of Sciences of the United States of America*, **111**: 7659–7664, 2014. doi: 10.1073/pnas.1322248111 (see pp. 43, 46, 66)

[94] S. Campagne, M. E. Marsh, G. Capitani, J. Vorholt, and F. H.-T. Allain. Structural basis for -10 promoter element melting by environmentally induced sigma factors. *Nature Structural & Molecular Biology*, **21**: 269–276, 2014. doi: 10.1038/nsmb.2777 (see pp. 43, 46, 90)

[95] H. Spieler. *Semiconductor Detector Systems*. Oxford scholarship online OUP Oxford, 2005. (see p. 48)

[96] DECTRIS *HYBRID PIXEL TECHNOLOGY & SINGLE-PHOTON COUNTING* 2012 (see pp. 48, 49, 51)

[97] C. PONCHUT. Characterization of X-ray area detectors for synchrotron beamlines. *Journal of Synchrotron Radiation*, **13**: 195–203, 2006. doi: 10.1107/S0909049505034278 (see p. 48)

[98] P. KRAFT. *PILATUS 2M*. PhD thesis. ETH Zürich, 2010. (see p. 48)

[99] M. MUELLER, M. WANG, and C. SCHULZE-BRIESE. Optimal fine ϕ-slicing for single-photon-counting pixel detectors. *Acta crystallographica Section D: Biological Crystallography*, **68**: 42–56, 2012. doi: 10.1107/S0907444911049833 (see pp. 49, 50, 55)

[100] O. B. ZELDIN, M. GERSTEL, and E. F. GARMAN. Optimizing the spatial distribution of dose in X-ray macromolecular crystallography. *Journal of Synchrotron Radiation*, **20**: 49–57, 2013. doi: 10.1107/S0909049512044706 (see p. 49)

[101] Z. DAUTER. Data-collection strategies. *Acta Crystallographica Section D: Biological Crystallography*, **55**: 1703–1717, 1999. doi: 10.1107/S0907444999008367 (see pp. 52, 53)

[102] O. B. ZELDIN, M. GERSTEL, and E. F. GARMAN. RADDOSE-3D : time- and space-resolved modelling of dose in macromolecular crystallography. *Journal of Applied Crystallography*, **46**: 1225–1230, 2013. doi: 10.1107/S0021889813011461 (see p. 56)

[103] Z. DAUTER, M. DAUTER, and E. J. DODSON. Jolly SAD. *Acta Crystallographica Section D: Biological Crystallography*, **58**: 494–506, 2002. doi: 10.1107/S090744490200118X (see p. 58)

[104] G. M. SHELDRICK. A short history of SHELX. *Acta Crystallographica Section A: Foundations of Crystallography*, **64**: 112–122, 2008. doi: 10.1107/S0108767307043930 (see pp. 58, 74, 88, 133)

[105] G. M. SHELDRICK. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta crystallographica Section D: Biological Crystallography*, **66**: 479–485, 2010. doi: 10.1107/S0907444909038360 (see pp. 58, 133)

[106] W. KABSCH. XDS. *Acta crystallographica. Section D: Biological Crystallography*, **66**: 125–132, 2010. doi: 10.1107/S0907444909047337 (see pp. 58, 133)

[107] PDB *Protein Data Bank Contents Guide* tech. rep. 2008, 205 (see p. 58)

[108] T. GRÜNE *Experimental Phasing with shelx c/d/e - Phasing in the Context of Structure Determination* tech. rep. 2013 (see p. 59)

[109] G. M. SHELDRICK. SHELXD for large small molecules and SHELXC / D for Macromolecular Substructures. 2007. (see p. 59)

[110] F. DALL'ANTONIA and T. R. SCHNEIDER. SITCOM : a program for comparing sites in macromolecular substructures. *Journal of Applied*

Crystallography, **39**: 618–619, 2006. doi: 10.1107/S0021889806014518 (see pp. 59, 133)

[111] K. Cowtan. DM. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography*, **31**: 34–38, 1994. (see pp. 59, 133)

[112] K. Cowtan and K. Y. J. Zhang. Density modification for macromolecular phase improvement. *Progress in Biophysics and Molecular Biology*, **72**: 245–270, 1999. doi: 10.1016/S0079-6107(99)00008-5 (see pp. 59, 76, 133)

[113] P. Emsley and K. Cowtan. Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D: Biological Crystallography*, **60**: 2126–2132, 2004. doi: 10.1107/S0907444904019158 (see pp. 60, 61, 104)

[114] P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, and P. H. Zwart. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica Section D: Biological Crystallography*, **66**: 213–221, 2010. doi: 10.1107/S0907444909052925 (see pp. 60, 66, 133)

[115] P. V. Afonine, R. W. Grosse-Kunstleve, V. B. Chen, J. J. Headd, N. W. Moriarty, J. S. Richardson, D. C. Richardson, A. Urzhumtsev, P. H. Zwart, and P. D. Adams. phenix.model_vs_data: a high-level tool for the calculation of crystallographic model and data statistics. *Journal of Applied Crystallography*, **43**: 669–676, 2010. doi: 10.1107/S0021889810015608 (see p. 60)

[116] A. J. McCoy. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallographica Section D: Biological Crystallography*, **63**: 32–41, 2007. doi: 10.1107/S0907444906045975 (see pp. 61, 103)

[117] Wolfram Research Inc. *Mathematica Edition: Version 9.0* Champaign, Illinois, 2012 (see pp. 61, 101, 133)

[118] N. Ambert, J. Vanwinsberghe, and P. Dumas. BioCrystallographica, a package for doing Crystallography with Mathematica. *CCP4 Newsletter*, **44**: 22–31, 2006. (see pp. 61, 101, 133)

[119] M. P. Robertson, Y.-I. Chi, and W. G. Scott. Solving novel RNA structures using only secondary structural fragments. *Methods*, **52**: 168–172, 2010. doi: 10.1016/j.ymeth.2010.06.011 (see pp. 62, 64, 104, 107, 123)

[120] F. JOSSINET, T. E. LUDWIG, and E. WESTHOF. Assemble: an interactive graphical tool to analyze and build RNA architectures at the 2D and 3D levels. *Bioinformatics*, **26**: 2057–2059, 2010. doi: 10.1093/bioinformatics/btq321 (see pp. 62, 112, 123, 133)

[121] C. C. CORRELL, J. BENEKEN, M. J. PLANTINGA, M. LUBBERS, and Y. CHAN. The common and the distinctive features of the bulged-G motif based on a 1.04 Å resolution RNA structure. *Nucleic Acids Research*, **31**: 6806–6818, 2003. doi: 10.1093/nar/gkg908 (see pp. 66, 103)

[122] P. V. AFONINE, R. W. GROSSE-KUNSTLEVE, and P. D. ADAMS. The Phenix refinement framework. *CCP4 Newsletter*, **42**: 1–18, 2005. (see p. 66)

[123] I. POLIKARPOV, A. TEPLYAKOV, and G. OLIVA. The ultimate wavelength for protein crystallography? *Acta Crystallographica Section D: Biological Crystallography*, **53**: 734–737, 1997. doi: 10.1107/S0907444997007233 (see p. 71)

[124] V. Y. LUNIN and M. M. WOOLFSON. Mean phase error and the map-correlation coefficient. *Acta Crystallographica Section D: Biological Crystallography*, **49**: 530–533, 1993. doi: 10.1107/S0907444993005852 (see p. 76)

[125] K. DIEDERICHS and P. A. KARPLUS. Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nature*, **4**: 269–275, 1997. doi: 10.1038/nsb0497-269 (see pp. 80, 144)

[126] M. G. ROSSMANN. The molecular replacement method. *Acta Crystallographica Section A: Foundations of Crystallography*, **46**: 73–82, 1990. doi: 10.1107/S0108767389009815 (see p. 99)

[127] M. DELARUE and P. DUMAS. On the use of low-frequency normal modes to enforce collective movements in refining macromolecular structural models. *Proceedings of the National Academy of Sciences of the United States of America*, **101**: 6957–62, 2004. doi: 10.1073/pnas.0400301101 (see p. 101)

[128] K. SUHRE and Y. H. SANEJOUAND. On the potential of normal-mode analysis for solving difficult molecular-replacement problems. *Acta Crystallographica Section D: Biological Crystallography*, **60**: 796–799, 2004. doi: 10.1107/S0907444904001982 (see p. 101)

[129] A. LAURENZI, L.-H. HUNG, and R. SAMUDRALA. Structure prediction of partial-length protein sequences. *International Journal of Molecular Sciences*, **14**: 14892–907, 2013. doi: 10.3390/ijms140714892 (see p. 102)

[130] G. SCAPIN. Molecular replacement then and now. *Acta crystallographica Section D, Biological crystallography*, **69**: 2266–75, 2013. doi: 10.1107/S0907444913011426 (see p. 102)

[131] E. KRISSINEL and K. HENRICK. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallographica. Section D: Biological Crystallography*, **60**: 2256–2268, 2004. doi: 10.1107/S0907444904026460 (see p. 115)

[132] BRUCKER-NONIUS INC. MADISON USA *XPREP* 2004 (see pp. 133, 135)

[133] G. M. SHELDRICK and T. R. SCHNEIDER. SHELXL: High-resolution refinement. *Methods in Enzymology*, **277**: 319–343, 1997. (see pp. 133, 135)

[134] J. KARLE and I. L. KARLE. The symbolic addition procedure for phase determination for centrosymmetric and non-centrosymmetric crystals. *Acta Crystallographica*, **21**: 849–859, 1966. doi: 10.1107/S0365110X66004079 (see p. 139)

[135] K. SHIN and P. J. HAMMOND. *Fundamentals of Signal Processing for Sound and Vibration Engineers*. Wiley, 2008. 244 (see p. 140)

[136] D. D. NOLTE. *Optical Interferometry for Biology and Medicine: Optical Interferometry for Biology and Medicine*. Bioanalysis Springer, 2011. 104–105 (see p. 140)

[137] D. A. MCQUARRIE. *Statistical Mechanics*. University Science Books, 2000. 554–558 (see p. 140)

[138] A. ROGERS. *Essentials of Optoelectronics with Applications*. Optical and quantum electronics series Taylor & Francis, 1997. 144 (see p. 140)

[139] E. W. WEISSTEIN "*Wiener-Khinchin Theorem.*" *From MathWorld–A Wolfram Web Resource*. 2014 (see p. 140)

[140] B. RUPP. *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology*. Garland Science, 2010. 556–557 (see p. 142)

[141] P. EVANS and A. MCCOY. An introduction to molecular replacement. *Acta Crystallographica Section D: Biological Crystallography*, **64**: 1–10, 2008. doi: 10.1107/S0907444907051554 (see p. 143)

[142] V. LUZZATI. Traitement statistique des erreurs dans la determination des structures cristallines. *Acta Crystallographica*, **5**: 802–810, 1952. doi: 10.1107/S0365110X52002161 (see p. 143)

[143] P. EVANS. Scaling and assessment of data quality. *Acta crystallographica. Section D, Biological crystallography*, **62**: 72–82, 2006. doi: 10.1107/S0907444905036693 (see p. 144)

Acknowledgements

I wish to express my sincere gratitude to everyone who supported me in succeeding this thesis. First, I would like to acknowledge the members of my PhD committee: Prof. Dr. Roland Sigel, Prof. Dr. Anthony Linden, Dr. Vincent Olieric, Dr. Meitian Wang and Dr. Philippe Dumas.

In particular, I would like to thank Prof. Dr. Roland Sigel for his willingness to supervise this project and the freedom granted during its completion.

My deep and special thanks go to my advisors Dr. Vincent Olieric, Dr. Philippe Dumas and Dr. Meitian Wang for their constant support and encouraging guidance throughout this thesis, and for their steady kindness and patience during my PhD. I would also like to thank Dr. Eric Ennifar, Dr. Joachim Schnabl, Dr. Mélanie Meyer and Michelle Schaffer for their kind grant of RNA crystals and all their effort for RNA purification and crystallization.

In addition, I am also very grateful to all the present and former members of the MX-group at Swiss Light Source of Paul Scherrer Institut for all the great times we spent together.

Finally, I would like to thank my family and friends for their unshakable encouraging and supporting during my PhD.

Curriculum Vitae

Surname	Peng
First name	Guanya
Date of Birth	05.03.1986
Nationality	People's Republic of China

Education

2011 – 2014

- PhD student at MX-group, Swiss Light Source, Paul Scherrer Institut, under supervision of Prof. Dr. Roland Sigel at Universität Zürich, Zürich, Switzerland
- Thesis title: On the use of phosphorus for the nucleic acid structure determination

2009 – 2010

- Institut polytechnique de Grenoble, Grenoble, France. Master Degree, Science et génie des matériaux (Erasmus Programme)
- Thesis title: Crystallographic structural studies of Al-Ta, Al-Nb-Ta and Al-Ta-V complex alloys

2008 – 2009

- Universität Augsburg, Augsburg, Germany. Master Degree, Advanced Functional Materials (Erasmus Programme)

2004 – 2008

- Harbin Institute of Technology, Harbin, China. Bachelor Degree, Optical Information Science and Technology
- Thesis title: Tm^{+3}, Er^{3+} co-doped Y_2O_3 white light fluorescent materials synthesis and fluorescence spectrum analysis

2001 – 2003

- The Second High School Attached to Beijing Normal University, Beijing, China. High School

Conferences Attended

2013

- 27th Rhine-Knee Regional Meeting on X-ray Crystallography of Macromolecules, Universität Freiburg, Germany, September 25-27 (20 minutes presentation)
- CCP4 Study Weekend 2013, Molecular Replacement, University of Nottingham, United Kingdom, January 3-5

2012

- Biostruct-X Workshop on Advanced Data Collection with Multi-Axis Goniometer and Single-Photon Counting Detector, Villigen PSI, Switzerland, November 6-12
- 26th Rhine-Knee Regional Meeting on X-ray Crystallography of Macromolecules, École polytechnique fédérale de Lausanne, Switzerland, September 26-28
- 4th Winter School on soft X-rays in Macromolecular Crystallography, European Synchrotron Radiation Facility, France, February 6-8
- CCP4 Study Weekend 2012, Data Collection and Processing, University of Warwick, United Kingdom, January 3-5

2011

- 25th Rhine-Knee Regional Meeting on X-ray Crystallography of Macromolecules, Sursee, Switzerland, September 28-30
- EMBO Workshop on Exploiting Anomalous Scattering in Macromolecular Structure Determination, European Synchrotron Radiation Facility, France, June 6-10 (poster)